

For Reference

NOT TO BE TAKEN FROM THIS ROOM

For Reference

NOT TO BE TAKEN FROM THIS ROOM

Ex LIBRIS
UNIVERSITATIS
ALBERTAEENSIS



THE UNIVERSITY OF ALBERTA

THE RELATIONSHIP OF DIFFERENT SCHEDULES
OF KNOWLEDGE OF RESULTS TO PERFORMANCE
ON PROGRAMED MATERIAL

by

HANS VON STACKELBERG



A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

APRIL, 1968



Digitized by the Internet Archive
in 2020 with funding from
University of Alberta Libraries

<https://archive.org/details/Stackelberg1968>

UNIVERSITY OF ALBERTA
FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommended to the Faculty of Graduate Studies for acceptance, a thesis entitled "The Relationship of Different Schedules of Knowledge of Results to Performance on Programed Material" submitted by Hans Von Stackelberg in partial fulfillment of the requirements for the degree of Master of Education.

ACKNOWLEDGEMENTS

The writer wishes to thank his supervisor, Dr. J. E. Bicknell, for his continuous guidance, his constructive comments and sincere interest. Also, the suggestions and the support of Dr. V.R. Nyberg and Dr. M.A. Nay are sincerely appreciated.

Thanks are extended to the Staff of the Ponoka Junior High School which allowed the writer to carry out his research with the Grade VIII classes. Furthermore, thanks are extended to Mrs. Linda Olsen for her cooperation in the typing of this thesis.

Finally, the study is dedicated to the writer's wife, Luise, for her confidence in him and her great patience and constant support.

This study was conducted under Project A, Alberta Advisory Committee on Educational Research.

ABSTRACT

This study was designed to test the hypothesis that in a linear small step program of the Skinnerian type effectiveness of learning does not depend on the amount of knowledge of results given by the programmer in the form of the correct answer. The writer further hypothesized that sex differences, differences in intelligence and in previous knowledge of the subject matter would have a greater effect upon the learner's success than would the amount of knowledge of results.

The experiment to test these hypotheses was conducted at the Ponoka Junior High School involving 122 Grade VIII students. A science program of 213 frames, Sun Earth Relationship, was modified to provide subjects with one of four schedules of knowledge of results: 100%, 50%, 25%, 12.5%. Each subject was assigned to one of the four types of program and was tested before and after having worked through the program on two parallel forms of a test supplied by the publishers. The criterion variable was the subject's performance on the posttest.

Linear Multiple Regression Analysis was used. This technique made use of vector notation and was well suited for computer operations.

The significant findings in this study were as follows: (1) The amount of knowledge of results had no effect on subject's ability to learn from this program; (2) The performance of boys did not differ from that of girls; (3) Only when previous knowledge of or familiarity with the material had been controlled did intelligence appear to have an influence on learning from this program.

TABLE OF CONTENTS

CHAPTER	PAGE
I. THE INTRODUCTION	1
The Problem	3
II. REVIEW OF RELATED LITERATURE	5
Definition of Terms	8
Programed Instructions	8
Linear Program	10
Branched Program	10
Adjunct Program	11
Knowledge of Results	13
Immediate and Delayed Reinforcement	13
Summary	24
Partial Reinforcement	27
Summary	34
III. THEORETICAL FRAMEWORK.	36
IV. THE PROCEDURE	47
Method and Experimental Design	47
The Sample	49
Procedures	49
Statistical Analysis	50
Description of Vectors Employed in This Study	52
V. RESULTS	55
Tests for Sampling Bias on the Control Factors .	55
Differences Among the Groups on Intelligence	
Test Scores	55

CHAPTER	PAGE
Differences Among the Groups on Previous	
Knowledge of Material	58
Differences Among the Groups on Posttest Scores .	60
Posttest Performance with Intelligence Controlled	62
Posttest Performance with Pretest Controlled . .	64
Posttest Performance with Control on Both	
Intelligence and Pretest Performance	67
Summary of Results of Analysis	68
Discussion	71
BIBLIOGRAPHY	74
APPENDIX	79

CHAPTER I

INTRODUCTION

Periodically, the educational scene is bombarded with the claims of a special interest group which insists that it has discovered "the solution" to the problems of the system. Generally these innovations run a regular course consisting of three phases: at the outset there is a bitter attack on current practice, accompanied by extravagant claims of effectiveness for the new departure; later, there are counter-attacks by traditionalists; still later, after the dust and outrage have calmed, less partisan individuals are usually able to sort out fact from fantasy and submit the new approach to careful scientific examination. The results almost always are less than its advocate claimed and more than its opponents granted.

Much has been written about programmed instruction over the last 12 years, most of it in the past 5 years. At first, there was a great emphasis on the gadget or the machine aspect of this new technology of education initiated by Skinner (1954). This was accompanied by widespread popular interest and often premature enthusiasm for the teaching machine. From the early 1960's on, some of this initial and often unfounded enthusiasm gave way to a larger interest in examination of this method. Where the initial emphasis on machines had obscured the more important features of the new technology, the interest shifted to the university labora-

tory and the underlying psychological principles have come under scrutiny. These principles form a recognizable theory that can be subject to hypothesis and testing.

The theoretical base from which Skinner developed this technology lies with the stimulus-response school developed by Thorndike, Guthrie, Watson and Hull. The movement today is not simply the mechanization of teaching nor has reinforcement theory of learning failed to go beyond the earlier stimulus-response formulations but, instead, the development of a new technology. Teaching efficiency has been shown to increase by two closely related measures: (1) the application of some demonstrably valid psychological principles derived from learning experiments and (2) the use of teaching aids and devices.

The new technology of education is the application of behavioral laws in modifying or controlling behavior. Such a technology became possible with the realization that we are actually referring to a verbal repertoire, controlled by the same laws as other behavior. (Holland, 1959, p. 34).

The effectiveness of programmed instruction in assisting the learner to acquire knowledge has now been well documented. It has been used to teach adults and children, superior and mentally retarded individuals. There seems to be no limit to the variety of material that can be programmed and then presented to the learner.

Current research is moving away from testing the effectiveness of programmed learning against that of "conventional instruction"

and is concentrating on the presentation variables, the response variables, and lately, also the learner as a variable, his personality characteristics.

THE PROBLEM

When Skinner applied his experience with animal training to human learning, he defined response-confirmation as reinforcement in the sense that it increases the probability that desired behaviors will appear at the proper time (Skinner, 1954). It is this reinforcement which the program provides immediately after the correct response has been made that is central to the Skinnerian theory of program instruction. This is one reason why most linear programs are written with very small step size, minimizing the error rate. However, experience with schedules of reinforcement in animal learning would suggest that the programmer would achieve greater resistance to extinction if learning took place on a variable-ratio schedule of reinforcement, rather than continuous reinforcement.

Therefore two questions must be raised: (1) why should knowledge of results be reinforcing? and, (2) can partial schedules of reinforcement be employed as a condition of training?

The research in the literature dealing with the above question is not clear. The experiments so far have not been too successful in identifying the incentive in programmed instruction that makes response-confirmation or knowledge of results (when correct) an act of reinforcement. There also have been a number of studies investigating different

schedules of knowledge of results in programmed instruction. Among these is a somewhat disproportionate number which report a finding of "no significant differences".

CHAPTER II

REVIEW OF RELATED LITERATURE

A review of the current literature on self-instructional devices can be divided into two main areas, those writers who have addressed themselves to the description of the new devices and their programs, and those who have attempted to answer questions that arise out of the use of this new technology. Most of the authors in the 1950's belong the first group. In the last few years much research has been carried on and could be grouped into three categories:

1. The learner
2. The task
3. The practice factors

This review will be concerned mainly with the last category, and specifically with the feedback factors involved in programs.

Over the last seven years a number of noteworthy reviews of the literature related to programmed instruction have appeared. While most writers have attempted to cover large aspects of this innovation in education it appears that the reviews of research in the area of reinforcement and related aspects are scarce. Krumboltz (1961) has reviewed a number of studies related to the use of prompts, response modes, immediacy vs. delay of reinforcement, schedules of reinforcement and error rate on programs. Many of the studies reviewed by Krumboltz are plagued by a number of factors which would tend to leave open to question any conclusions drawn from them.

Characteristically, the experimenters have used very small groups of subjects, Melaragno (1960), Holland and Porter (1960) have failed to report on the validity of their tests or have used extremely short programs and short criterion tests. Krumboltz found that the research was far from unequivocal and thus no clear picture emerges.

In a recent review of related literature, Stembitsky (1964) reviewed five studies investigating the effect of immediate vs delay in confirmation and eleven studies investigating the effectiveness of feedback vs no feedback. Some of these studies will be reviewed below. However, it is of interest that in most instances Stembitsky (1964) found it necessary to rate the studies as rather unreliable and inconclusive. He concludes that, "No specific findings were rated as highly dependable. From the overall point of view, it was found that the indices supported the conclusion that immediate confirmation was as good or better than delayed confirmation." (p. 117). One of the earlier surveys into the field of programmed instruction is the source book edited by Lumsdaine and Glaser (1961). The book is largely a compilation of research reports along with the articles not based on experimental evidence. It does not attempt any objective evaluation of the conclusions and procedures reported but some of the specific articles will be reviewed below.

Wilbur Schram (1962) published a small booklet in which he discussed the field of programmed instruction with reference to research. His findings indicate that much more research with improved procedures is needed in order to have more authoritative answers to some of the questions now being asked about programmed

instruction. He followed this booklet with an annotated bibliography of research, reviewing 190 articles and theses. Schram reports that there is a somewhat disproportionate number of experiments which report a finding of "no significant differences".

The most critical review of the work of designers of immediate reinforcer teaching machines comes from Porter (1957). It is also the oldest, and therefore not as complete as later reviews. However, it is still important because Porter has charged that certain assumptions made by the developers of teaching machines are not too well documented. He summarizes his criticism under the following headings:

1. generalization of results from animal studies to homo sapiens,
2. control of conditions which made knowledge of results reinforcing,
3. type of reinforcement used,
4. quantity of reinforcement,
5. schedules of reinforcement (Porter, 1957, p. 128).

In the following review research findings related to the above criticisms raised by Porter will be examined in the light of more recent findings in programmed instruction. It will be presented under the following headings which appear to represent useful limits.

1. Definition of terms essential to the understanding of programmed instruction.
2. Immediate and Delayed Reinforcement.
3. Partial Reinforcement.

DEFINITION OF TERMS

Programed Instructions:

Automated instruction, programed learning and programed instruction are to be considered synonymous and can be broadly used to include any method whereby the teacher's function is provided by a wholly or partially automated sequence of instructional segments prepared in advance of the learning process. This sequence is most often referred to as a program. (The spelling of program and programed with one "M" instead of two has become a common practice among program writers to indicate that it refers to this new development in education.)

Wilbur Schram (1962, p. 1) speaks of programed instruction as "the kind of learning experience in which a program takes the place of a tutor for the student, and leads him through a set of specified behaviors designed and sequenced to make it more probable that he will behave in a given desired way in the future...."

H.J.A. Goodman (1962, p. 226) advances a definition of programed instruction in which he described it as "...a strategy for presenting information to an individual and to develop his understanding step by step. It is a special strategy, however, because it maintains the best conditions for learning."

A number of reviewers (Schram, 1964; Cohen, 1962; Stolurow, 1962) have pointed out the similarities between the teaching method of Socrates and the principles of programed learning.

Socrates demonstrated the effectiveness of a logical, continuous interaction between teacher and pupil. In the method of dialectic teaching as in programmed instruction every question posed is intended to lead the learner step by step, in a logical, coherent sequence toward greater understanding. It may well be conceded that the methods on which programmed instruction are based are indeed not new (Cohen, 1962; Stolurow, 1962; Fusco, 1962). The newness, according to Fusco (1962), is the development of a technology based on a demonstrated principle. Through the use of this new technology it has become possible to provide a large number of children with individual self-instruction not unlike the Socratic question and answer method.

Another important principle underlies this new technology - that of active and continuous student participation. Elley (1963, p. 12) in a discussion of programmed learning states that "numerous studies have shown the superiority of active recitation and continual self-testing over the more passive process of rereading a verbal list or passage ...(and)...that Skinner (too) insists on active learners."

Essentially then, it is the program that teaches and involves the learner actively while at the same time allowing him to proceed at his own rate, thereby providing for individual differences in learning ability. These programs usually consist of a series of items of information constructed in a logical progression which requires the learner to respond either covertly or overtly. By responding in the appropriate manner the student

demonstrates the mastery of knowledge previously acquired through interaction with the program. This information is then overlearned in further use to build up a store of knowledge.

Linear Program:

The linear program is most associated with the theory and techniques of B.F. Skinner. This program is distinguished by its fixed sequence of items (frames) and requires all subjects to proceed through the same number of frames. It usually requires the subject to respond overtly (write answers to questions) and tends to minimize step size (see below). Because this type of program was used in this study further discussion of its merits and shortcomings will follow in the chapter on the Theoretical Framework.

Branched Program:

This type of program was first developed by Norman Crowder and allows for greater adaptation to individual differences than the linear program. In a branched program the learner's response determines the sequence of frames through which he will proceed. An incorrect response to a question will lead the subject to be "re-routed" to sets of items designed to correct his erroneous thinking. A fast learner on the other hand will be allowed to proceed as rapidly as he demonstrates the acquisition of material gained through using the program. Responses in a branched program usually take the form of selecting the correct answer in a multiple choice question at the end of a block of information.

Adjunct Program:

This type of program, with its testual material followed by a quiz provided only little opportunity for covert learner-program interaction. Important aspects of what is to be learned are further clarified and emphasized, but test items according to Pressey (1964), the main advocate of adjunct programing, should be on the multiple choice variety.

Response:

There are two categories of responses elicited from the learner in programed instruction.

Covert Response: This type of response only requires that the subject think about an answer to a frame, no further activity on his part is required.

Overt Response: A physical manifestation on the part of the learner is required to constitute an overt response to the stimulus presented in the form of a question. This usually takes the form of writing an answer, verbalizing, pushing a button or punching a hole in a card. This important aspect seems to be that the learner has to act and to make a decision and commit himself to a response.

Frame:

Ordinarily a single item of information, the smallest single unit of the program, is designated as a frame in the Skinner type

program. It constitutes a set of stimuli, capable of eliciting a response from the learner. Recent developments in programing, however, have made necessary a redefinition of this term and it is now used to include any material presented to the subject at one time.

Error Rate:

This term is used to refer either to the proportion of errors made by a specific student or group of students on a total program, or the percentage of a given group responding incorrectly to a specific item in the program. The latter is the criterion usually used to determine revision in the program. In a later discussion the role of error rate to the basic method of programing will be discussed.

Step Size:

This is the measure of the increment in difficulty level between steps of a program. Large steps require greater effort from the learner and usually result in an increase in error rate. Step size may vary greatly for different programs depending upon for whom the program has been written. If it is new material and the learner is young or dull, step size will be at a minimum, on the other hand, if mature learners are involved and they already have a certain familiarity with the material step size will be increased, resulting in a more difficult but shorter program.

Prompting or Cueing:

A cue or prompt refers to a deliberate and purposeful part of the program presentation to make the correct response more likely.

Knowledge of Results (K of R):

One of the fundamental principles of program instruction theory is the necessity to provide the learner with a constant opportunity to judge his progress by comparing his responses with those provided in the program. This may be by means of some auditory or visual stimulus or in the case of constructed answer by requiring a comparison between his answer and that of the programmer. There is, of course, another way in which knowledge of results is possible. In a well constructed program of the linear type, each frame depends to some extent on the preceding one, often also concepts are repeated in different contents and with different examples. It is therefore quite impossible for the learner to judge the correctness of previous responses by implicit knowledge of results. Further discussion of this important concept will be found in the theoretical chapter.

IMMEDIATE AND DELAYED REINFORCEMENT

Skinner has pointed out in an article appearing in Science 1958 that reinforcement in the form of knowledge of the correct response should come immediately after the response has been made. He sees in the teaching machine the ideal dispenser of immediate and continuous reinforcement. The following studies examined the effect of delay of reward on learning.

Terrell and Ware (1961) investigated the effectiveness of an immediate reward as compared with a delayed reward in the speed of size and form discrimination learning in childhood. The reward

employed was a light flash. Forty children in elementary school were used in the experiment. All stimuli were three dimensional geometric figures. There were two categories each of two experimental variables. One variable was the reward condition while the other was the task factor. The two reward categories were (a) an immediate light flash following a correct response and (b) a seven-second delay in the light flash following a correct response. The two tasks were (a) a button-pushing response to the larger of two cube boxes and (b) a button-pushing response to a sphere instead of a pyramid. Subjects were randomly assigned to two groups of twenty children each. Group I was given the size task under the delayed reward condition and the form task under immediate reward condition. Group II was given the reverse of the above.

Results:

Subjects learned significantly more quickly when rewarded immediately than when given a delayed reward. The results were significant at $p. = .0001$.

The question of why an immediate reward should result in more efficient discrimination learning than a delayed reward must be dealt with. Is the associative factor being affected by time interval between response and reinforcement or is it the incentive factor? Both may be affected. Subjects responded quicker and made choices quicker under immediate reinforcement than when learning under delayed conditions. This may suggest that subjects were more certain of their choices being correct ones when rewarded immediately. On the other hand, the subjects under delayed reward conditions

exhibited much less concentration which showed itself in such activity as looking out of the window and asking irrelevant questions.

Perin (1943) investigated the relative effects of immediate and delayed reinforcement in associative and incentive factors in two-stimulus size and form discrimination learning involving children. Reward consisted of a light that went on as soon as the subjects pressed a button for the correct answer. For one task the correct response was a button-pushing response to the larger of the two cube boxes and for task two a response to the sphere instead of the pyramid. It was found that subjects learned significantly more quickly under immediate than under delayed reinforcement conditions ($p = .005$). This study shows delay in reward may thwart problem solving behavior. This supports the conclusion of the previous study by Terrel and Ware (1961). Another possible explanation for these results may be that subjects became uninterested and consequently bored with a task which involved delay of reinforcement and so the learning efficiency decreased appreciably.

In another study Perin (1943) used 100 albino rats to investigate the effect of delayed reinforcement. The rats were divided into five groups and trained on a bar-pressing response. Group I received the food pellet reward simultaneously with the occurrence of the response; Group II received food after a delay of two seconds following the bar response. Group III was given food after five seconds' delay and Group IV after 10 seconds' delay. A fifth group made up of 16 animals received training with food reinforcement being delayed 30 seconds following the bar response. Fifty reinforced trials were given to the animals in each group.

The measure of habit strength was response latency (the time required for the response to occur after an animal had been given access to the bar). The rate at which the practice curves decreased showed a consistent relation to the length of the delay period. Animals reinforced simultaneously with the response demonstrated the most rapid learning, with the rate of learning becoming increasingly slow as the delay interval was increased. Two equations were given which described the exponential constants for the practice curves as functions of delay interval. Indications were that within the range of 14 - 21 seconds' delay the exponential constant probably reaches zero. This was considered to indicate that on the average learning cannot take place with a delay interval larger than 14 - 21 seconds. Perin was very careful to eliminate any conditions that may have acquired properties of secondary reinforcers.

In a third experiment Perin (1943) used six groups of 25 albino rats, which were trained in a bar-pressing situation. The bar was capable of being moved in a right or left position and the task demands of the animals was the selection of one of these directions of push as the correct response. Provision was made for the bar to withdraw automatically following a correct response. The experimental variable differentiating the six groups was the interval of time between the correct response and food reinforcement for the response. Delay intervals used, corresponding to the six groups, were zero, 2 seconds, 5 seconds, 10 seconds, 20 seconds, and 30 seconds. All animals were given 60 training trials on the bar responses with reward being given immediately for either direction

of push. Differential reinforcement was then given to animals with the preferred response being incorrect and with the alternative response being reinforced following delay intervals for the experimental group. A correction technique was employed by which each trial ended with a correct response. Practice curves for five of the six experimental groups showed the percent of trials made without error. The limits of habit strength in terms of perfection of the response differentiation with practice did not discriminate between zero and two-second delay groups. Additional delays (5, 10 and 20 seconds) resulted in a decrease in the level of perfection.

The greater the delay of reinforcement the less the rate at which the habit was acquired. The results indicated that if food was delayed longer than 30 seconds the correct response would not be learned. If reinforcement is delayed beyond 2 seconds the intervening variable of secondary reinforcement may have affected the results. The author found it difficult to eliminate secondary reinforcement characteristics but in the highly controlled experiment kept this down to a minimum. However, this shows that under other conditions secondary reinforcement may be a very significant variable in learning if the reward is not immediately forthcoming.

In a study carried out in 1949 Angell tried to determine the effect of immediate and delayed knowledge of quiz results on three types of learning in freshman chemistry. The types of learning were knowledge of facts and principles, applications of facts and principles in non-quantitative problems and application of facts and principles in quantitative problems. Students in the experimental group obtained immediate knowledge of results by means of the Angell

and Troyer punchboard. Delayed knowledge of results was obtained through the use of I.B.M. answer sheets which were scored and returned to the students at the meeting following the taking of the quiz, an interval of about 24 hours. Differences between scores on the final examination were statistically significant in favour of the experimental group that used the punchboard and received immediate knowledge of results.

This study tends to support the findings of others that knowledge of results has to be immediate in order to be most effective. But since both groups improved over the period of instruction immediate knowledge of results cannot be said to be necessary for learning. It is regrettable that the experimenters did not separate the scores on the basis of the type of learning involved, e.g., knowledge of facts and principles, application of knowledge of facts and principles in non-quantitative problems, application of facts and principles to quantitative problems, neither did they administer a retention test after a 3 or 6 week period.

Another interesting point emerges here. Because the material was not a program of the Skinnerian type but rather in paragraph form with a quiz following each period's study, immediate knowledge of results in the quiz may have had a more important role than it would have had on a program which was structured in sequential order with small steps between individual frames. In such a program intrinsic knowledge of results may be obtained by the learner independent from that provided by the programmer.

In an ambitious study involving 220 university students enrolled

in introductory psychology the experimenters Moore and Smith (1961) attempted to compare several types of immediate reinforcement. They used 1,152 frames of the Holland-Skinner psychology program. It was divided into 12 units one of which was completed in each laboratory session. For some of the experimental groups slight modifications were made in the program.

The subjects were divided into 10 groups, five used a multiple-choice program, five a write-in type program. Four out of each of the five groups used teaching machines, the two remaining groups were given programed texts. The first pair of groups were not given knowledge of results. The second pair of groups were told immediately what the correct response should be. The next two groups were told immediately if the response was correct. The fourth of the two groups were given immediate knowledge of results plus an extrinsic reward in the form of one penny for each correct answer. The fifth and last pair of groups worked from programed text books and immediate knowledge of what the correct response should be was provided. No other formal instruction in learning theory was provided for the subjects during the experimental period. Two examinations were administered, one at the end of the first six units and one at the end of the experimental period covering all concepts presented in the experimental period.

Results:

An analysis of variance was applied to the data to compare the differences of means resulting from the type of knowledge of results provided, the method of responding and the interaction of these variables. None of the F's reached the necessary level of significance. Therefore,

the experimenters concluded that any differences of groups can be attributed to chance. There are several possible reasons why the different treatments have failed to differentiate. Learning efficiency may have been affected by the implicit reinforcement which was given within the frame through the use of extremely small steps and further confirmed through successive frames. The program is comprised of extremely small steps with considerable redundancy. Also, the experimenters unfortunately did not administer a pretest, thus there is no indication as to the amount of previous knowledge the students may have had in this area. Further, the tests are not described and may have lacked in sensitivity to bring out significant differences between methods.

Saltzman (1951) made two attempts to determine the effect of delaying reward on speed of human learning. The learning task was a verbal maze and information that a response was correct was the reward. Forty-four college students served as subjects in each experiment. There were two groups of students (a) delayed reward, and (b) immediate reward. The first experiment, although it provided no information concerning the effects of delayed rewards on learning, did provide information with regard to the speed of presenting learning material and the rate of learning. Reducing the presentation time from 18 seconds between items to 12 seconds did not affect learning rate significantly. The reward was not effectively delayed since students could learn the pattern of responses without learning the task they were assigned to. In the second study the numbers to be learned appeared in different order on successive trials thus the left right

position could not be memorized and the numbers themselves had to be learned. In this study the reward was delayed 6 seconds and it was found that the delayed reward group required 14.09 trials to criteria while the immediate reward group required only 9.95 trials; the difference is significant at the .01 level of confidence. The author assumed that knowledge of reward is reinforcing to the students. It must be pointed out too that informational feedback had to be delayed also, however, this study seems to support the view that for human subjects knowledge of a correct response is reinforcing.

Logan (1952) set out using 40 rats which were trained to respond to two bars, one followed by reinforcement after one second and the other after five seconds. The subjects showed significantly greater response speed to the short delay bar and developed a significant preference for this bar. After 80 trials the delays of reinforcement were reversed for half of the subjects. An appropriate reversal in the response speeds and preference followed this change, while no such change occurred in the control group while it was continued under original delay conditions.

Hull and Logan's work introduced a new aspect to the problem of how important is knowledge of results to effective learning. One aspect of their work would support the assumption that knowledge of results has to be immediate to be most effective, the other aspect dealing with the magnitude of the reinforcer is much more complicated when applied to programmed instruction, and is not dealt with in the review.

In a study investigating a number of variables of which

immediate knowledge of results was one, Evans, Glaser and Homme (1962) found that delay in feedback from 30 seconds to 5 minutes had no significant effect on learning symbolic logic as measured by an immediate posttest and three retention tests at one week intervals.

Feldhusen and Birt (1962) studied a number of variables in an experiment involving thirty college students. They found no significant difference in the performance of subjects who worked through a program on teaching machines in which all blanks were filled in from one where all feedback took place at the end of the program or where none was provided. Because of the small number of subjects and large number of variables their results have to be received with caution. Another weakness of this study is the small number of frames (37) used in the program, and the shortness of the posttest. Furthermore, this study was carried on with superior adults and thus the absence of the feedback or its delay until the program was completed may not have had the effect that similar procedures would have if a longer, more difficult program had been used with less sophisticated subjects.

The findings of Meyer (1960), however, seem to support the hypothesis that immediate feedback is superior to delayed feedback in producing learning. Her study involved three experimental groups. Group I received no feedback on the Grade VIII Language program. The answers of the subjects were checked by the teacher and the booklets returned the next day. Group II subjects scored their own answers by putting an "X" on wrong answers. Group III subjects were instructed to return to missed items and correct them. The experimenter found that greater learning occurred in subjects in Groups II and III where the

feedback was not delayed. However, Meyer appeared to have questioned her own findings on the grounds that there existed a lack of equivalence in the initial distribution on the pretest.

In two further experiments using auto-instructional materials Moore and Smith (1961) tested the hypothesis that information of the correctness of each response must be provided to the student with a minimal delay for effective learning to take place. Both experiments used Grade VI students in coeducation classes and with an IQ spread from about 70 to 125. The material to be learned consisted of spelling words. Subjects were pretested and posttested with the Metropolitan Spelling Achievement Test, form R.

In the first experiment subjects were divided into two groups. The Experimental group used a small answer booklet from which subjects could obtain the correct response to a frame within a few seconds after writing the response. The Control group used the same program but without the answer booklet. Subjects were tested each week after having worked on the program for 15 minutes on the first three days of the week. Moore and Smith did not find the experimental group to be superior on either the weekly tests or the final test. They concluded that their findings do not support the assumption that knowledge of results should be immediate. However, because their experimental design included some aspects which may have confounded the findings, they improved their design in the second experiment. First, the number of words to be learned was reduced from 400 to the 210 words missed most frequently on a pretest, and the number of words studied in one week was increased from 17 to 35. Second, the Metropolitan

Spelling Test was dropped as the criterion and a posttest of forty words chosen at random from the 210 was substituted. Third, the programed material was presented in the Foringer Teaching Machine. For the experimental group the correct answer appeared in the window of the machine if the lever was pulled, for the control group the window was masked preventing subjects from receiving knowledge of results through the machine. Again the experimental group failed to achieve significantly higher scores than the control group. On some of the weekly tests the control group scored higher than the experimental group but the results failed to reach significance at the .05 level.

The authors rejected the hypothesis that immediate knowledge of results is crucial to effective learning. They go on to explain that it may be possible that knowledge of results in the form of the correct response may be a form of overcuing, especially if the correct response on an item before may provide a prompt in the next frame. They concluded that the hypothesized beneficial aspect of knowledge of results in the form used in their program may in fact reduce learning efficiency when the correct answer serves as a cue to the next item. It may be that overcuing causes the student to become bored with his task. An alternative to this situation in the opinion of this writer seems to be the reduction of cues and prompts in a program which would make use of frequent knowledge of results.

SUMMARY

From the research reviewed here certain conclusions can be drawn. It appears that reinforcement in the form of knowledge of

results as provided by the program is not an essential feature in learning. The studies involving animals are well documented and the findings support the hypothesis that delay of reward has a gradient effect on the rate and progress of learning. Early studies with animals have established gradients which fell off rapidly at first but extended well beyond one minute, however, the studies reported by Perin (1943) have reported much shorter reinforcement gradients. Extrapolation of the curve fitted to one set of Perin's data indicated that the effectiveness of the reward fell to zero at a delay between 30 and 40 seconds while the second study suggested the possibility of an even shorter gradient.

In analyzing the various experimental results one is led to conclude that a major factor operating in the delayed reward situation is the degree to which conditions are favorable for secondary reinforcement. This observation was made also by Hull. Hull (1943) explains the failure of two early studies, one by Watson, G.B., and one by Warden and Hass (1934) to find any difference in the rate of learning of a delayed and a non-delayed group as due to the fact that the animals were admitted immediately into the feeding compartment following the correct response, although they were prevented from eating for the delay period. The goal box and the odor from the food container, however, provided immediate secondary reinforcement as these cues had acquired this property either from the preliminary training or previous experience with the sight and odor of food. In other studies in which the delay occurred in a separate detention compartment outside of the food box, secondary reinforcement cues of the goal box were also delayed.

However, the stimulus cues in the delay compartment also came to acquire secondary reinforcing properties because of their temporal and spatial association with the goal box and food.

Perin's studies, which were carried out as a part of a coordinated research program under Hull's direction, were specifically designed with a view to eliminating secondary reinforcement. He used a single compartment with three functions: (1) as a problem box, (2) as a delay chamber and (3) as a food box. Perin writes concerning this procedure: "the stimulus environment of the animal would be much more constant during the delay interval, and the train of external stimuli leading to the food reward characteristic of the maze situation would be eliminated; this would materially simplify the problem of interpretation by eliminating an irrelevant factor" (1943, p. 38).

Perin's results seem to bear out his explanation. Both experiments indicated that learning would not occur if the food were delayed longer than 30 - 40 seconds.

However, in the area of human learning and learning from programmed instruction the evidence is not as conclusive. It appears that where experimenters have used linear, small step programs, reinforcement in the form of knowledge of results has had a very small or negligible effect. It seems unfortunate that the area of knowledge of results as a reinforcer is confounded by the possibility that intrinsic or incidental knowledge of results may not have been controlled effectively in experimentation. How far this intrinsic knowledge of results can be viewed in terms of a delayed reinforcer and to what extent this delay has influenced learning negatively is

also difficult to determine. Delay, however, may be important.

PARTIAL REINFORCEMENT

Little more than twenty years ago the majority of psychologists interested in learning theory first became aware that there are important differences between continuous and non-continuous reinforcement, particularly when resistance to experimental extinction is used as a measure of response strength. Various animals, including humans, were reinforced on some but less than 100 percent of a series of trials or responses in situations where many kinds of behavior were measured.

For a long time response strength appeared to be proportional only to the number of reinforcements without consideration of the percentage of responses rewarded or the pattern of reinforced and non-reinforced responses (Hull, 1943; Perin, 1942; and Williams, 1938). Experimenters assumed that fewer reinforcements and partial reinforcements would produce less response strength than a larger number of reinforcements. Later evidence has not substantiated this postulate. A wave of partial reinforcement studies appeared in the literature around 1939-40, involving many kinds of behavior -- eye blink, leg flexion, salivation, GSR -- that dealt with the effects of this variable on acquisition, maintenance of behavior and resistance to extinction.

Lewis and Harder (1959) investigated the effect of partial reinforcement on retention of learning after a long series of continuous reinforcement acquisition. Two groups of 10 rats were given 100 continuously reinforced acquisition trials, bringing response strength virtually to an asymptote. Then one group remained on a continuously

reinforced schedule for 30 trials and the other was shifted to a partial schedule for 30 trials. Both groups were then extinguished. The results indicated that: (1) the onset of partial reinforcement had at least an initial depressant effect; (2) partial reinforcement, even after a long continuously reinforced series, resulted in increased resistance to extinction; (3) partial reinforcement has its effect on a performance variable and not upon increasing learning. This study appears to have some implication to program instruction as it would agree with Skinner's findings that reinforcement should be continuous but in humans 100 percent reinforcement may not be required and could even be detrimental to learning efficiency because it would tend to become monotonous.

An ultra-simple theory of reinforcement could be put in the following terms: Each time a conditioned response is reinforced, its strength is increased; and each time it occurs without reinforcement its strength is diminished. Such a simple theory can be easily put to the test. Instead of giving reinforcement on every trial during the conditioning series, intersperse unreinforced trial and finally proceed to an extinction series in which every trial goes unreinforced. Extinction should be very rapid, being indeed already partly accomplished by the preceding unreinforced trials.

In the bar-pressing task Skinner (1938) found it possible to maintain a high rate of response by giving reinforcement on only a part of the trials. If the rat, after being conditioned to press the bar and obtain a pellet, presses but obtains no pellet he is likely to press again before giving up; and if after several unrewarded pressings

he does obtain another pellet, he keeps up the pressing activity. After such intermittent reinforcement (either "variable" with a pellet say once a minute; or in "fixed ratio" with a pellet after 50 pressings) an extinction series requires a great number of trials.

Skinner and his associates have been making an extensive investigation of the effects of various schedules of reinforcement on the rate of responding, the results have been published in book form (Skinner and Fester, 1954). These studies have used pigeons rather than rats. It has been found they maintain surprisingly high rates of responding with infrequent rewards; for example, with only one reinforcement every 10 minutes, the pigeon may peck 25 times per minute for the whole 10 minute interval (Skinner, 1953). The response was usually a peck at a button or key -- a "natural" response for the pigeon.

Humphreys (1938) tested the hypothesis that after intermittent reinforcement there should be no sharp change at the beginning of the extinction series and extinction should be delayed and gradual. Humphreys used two groups each of 22 students; one group was always given reinforcement, the other group was given reinforcement 50 percent of the time; the conditional response was the eyelid response after a puff of air and the flashing of a light. The 100 percent group got the puff on every trial, the 50 percent group on only half the trials in random arrangements. A third group were given blank trials (rests) in place of the unreinforced trials of the 50 percent group, and gave the same results as the 100 percent group. Results: the interspersed unreinforced trials slowed the acquisition of the CR very slightly, but

the frequency level reached in 96 trials was nearly the same after 50 percent and 100 percent of reinforced trials. So the extinction series started from approximately the same level in both groups. But the first three extinction trials produced a sudden drop after 100 percent reinforcement, and no immediate drop at all after 50 percent reinforcement.

The same results -- slow or delayed extinction after intermittent reinforcement -- was obtained by the same author (Humphreys, 1940) with the human G.S.R. This "Humphrey's effect" has later been verified by other experimenters and in several learning tasks such as the simple runway and T-maze with rats.

One of the classical studies in this area has been reported by Skinner (1938). After previous conditioning and participation in other experimental activities, four rats were run in a rotated order under several different periodic reinforcement regimes; three, five, seven and nine minutes. Each rat underwent two one-hour periods on successive days at each interval. The mean number of responses per hour were for the four periodic intervals respectively: 319, 266, 220 and 160. The relationship is essentially linear over the range of four points involved.

A second kind of partial schedule, called fixed ratio reinforcement, was also developed and studied by Skinner (1938). The operations defining it consist of presenting a reward for a fixed number of responses. Thus, a rat might be reinforced after every twentieth bar-pressing. Skinner's basic experiment consisted of running six previously conditioned rats through a random order of fixed ratios: 48, 96 or 192

responses for one reinforcement. The characteristics of the curve are very interesting.

- (1) Just after receiving food for a response there is a pause that increases as the ratio decreases.
- (2) There follows a period of gradually increasing responding that
- (3) seems to reach a stable state shortly before the next reinforcement.

It should be noted that the behavioral effects of variable and fixed ratio reinforcements are opposed. In the first procedure increasing the time interval between reinforcements produces a decreased rate of responding. Under fixed ratio reinforcement, increasing the number of responses per reinforcement tends to yield a higher rate overall. Skinner's (1938) cumulative response curves under periodic reinforcement are smooth, lacking the discrete, step-wise nature of the fixed ratio curves. Thus, one can say fixed ratio reinforcement strengthens rapid responding; periodic reward ultimately strengthens responding. The earliest extinction comparison of periodic and continuous reinforcement has been reported by Skinner (1938). Although not providing numerical information, he notes that the post-periodic cumulative extinction curve in the bar-pressing situation as compared with the post-continuous one rises more slowly, reaches higher asymptote, falls off more slowly and lacks the cyclic fluctuations which are attributed to emotional effects in the continuous case. He concludes that "no amount of continuous reconditioning will yield an extinction curve of the height

obtained through even small amounts of periodic reconditioning." (p. 138)

Heron (1942) used a two bar situation and gave 12 days of extinction to one-half of the rats on the right lever and to the other half on the left lever. He used 48 rats which he had trained to operate the two levers, he set the right lever to "pay off" a pellet every three minutes and the left one each six minutes. Under these conditions the rats began to prefer the right lever. Responses in extinction were more frequent on the formerly three minute reinforcement lever for the first four days. On the sixth day the average number of responses on the two levers used was about the same. On days 7 - 12 more responses were given on the lever formerly producing reinforcement once every six minutes. These findings are consistent with the theory that the low rate for low frequency of reinforcement in conditioning is maintained in post-periodic extinction so that if the latter is prolonged the ultimate number of responses exceeds that for the lever more frequently rewarded in training.

Using 121 students Krumboltz and Weisman (1962) tested the effect of intermittent confirmation in a 177 frame programed textbook. They used 100% knowledge of results, two fixed ratios of knowledge of results of 67% and 33%, two variable ratios with the same percentage and one group that received no knowledge. The authors describe the pre and posttest and report split-half reliabilities of .85. Answers to the program were to be recorded by the students. The authors failed to find significant differences of the mean scores on the criterion test between the 6 groups. Thus they conclude that it does not seem to matter if there is knowledge of results, nor that there is any one method

of dispensing it to maximize its effectiveness. However, there was found a significant negative relationship between the amount of confirmation (knowledge of results) and the number of errors made on the program itself; the more confirmation, the fewer errors. It may be hypothesized that the effects of various schedules should become increasingly evident as more of the internal cueing is removed from the program.

One of the questions to be raised with this study is the lack of a retention test after some time has elapsed. An immediate posttest may lack some of the sensitivity that would have accounted for differences of learning with different schedules of knowledge of results. Weisman (1962) found that a retention test delayed two weeks was more sensitive to the effect of overt responding in an experiment involving modes of responding.

The findings of Krumboltz and Weisman (1962) seem to confirm such a hypothesis which could also be supported by some earlier evidence of Evans, Glaser, and Homme (1960) that the relevance of variables such as response mode and immediacy of confirmation is inversely related to the probability of correct responding but may not be so related to learning as measured by posttest.

In a M.A. thesis by Eugenia S. Scharf on the effects of partial reinforcement in a learning program the program was rewritten to allow for four treatments of knowledge of results, 100 percent continuous reinforcement, 50 percent fixed ratio and 50 percent variable ratio reinforcement and 25 percent variable reinforcement. Reinforcement was the correct answer. She found no significant differences on a multiple-choice criterion test.

Schram (1964) suggests that results of this nature can perhaps be explained by showing that reinforcement by confirmation is a crucial event only when probability of correct response is low, but in Skinnerian programs the probability is deliberately kept high (p. 53).

SUMMARY

The theoretical and practical significance of partial reinforcement cannot be overestimated. Any comprehensive system of behavioral principles must account for the fact that behavior can be established and maintained with reward occurring only a portion of the time. The findings regarding resistance to extinction after partial as contrasted with continuous reinforcement must be included in this treatment.

Partial reinforcement seems to be a basic condition of everyday behaving, rarely does the presentation of reward follow each time a response occurs. Continuous reinforcement seems to be the exception, not the rule. Both practically and experimentally, partial reinforcement is a worthwhile procedure. It retards satiation effects and permits continued effort over long periods and large number of trials. The evidence reviewed here suggests the efficacy of partial reinforcement in everyday life. Learning is not greatly retarded by its use, though it appears advantageous to use 100 percent reinforcement in conditioning simple responses initially. The behavior is more than adequately maintained with partial or infrequent reward. Partial reinforcement through knowledge of the correct response has had no negative effect on learning from programmed instruction (Scharf, 1962; Krumboltz and Weisman, 1962).

Finally the most striking aspect of the data is the far greater

resistance to extinction found after variable reinforcement as compared with continuous. If one desires behavior to be maintained for long periods of time in the absence of externally presented reward, variable reinforcement should be used in maintenance training. It appears that stimulus-response principles provide a framework accounting for the phenomena in this area.

CHAPTER III

THEORETICAL FRAMEWORK

The striking discrepancies found between the findings of researchers in the importance of knowledge of results as provided by the program appears to call for a close look at the learning process that takes place when a subject is exposed to programmed material.

Skinner (1954) has based his theoretical position on programmed instruction on his experimental work in the area of free operant conditioning. The subjects are faced with a stimulus S_1 , to which there is one correct answer R when the response is made it terminates S_1 and brings S_2 a reinforcing stimulus. This is usually provided by the programmer in the form of knowledge of the correct response. Zeaman (1959) has argued that while some of the literature of free operant conditioning is relevant the literature of controlled operant is even more important.

Zeaman points out that since in a programmed instruction situation only one answer will represent the correct response and in that the response will terminate the stimulus situation it has more of the characteristics of controlled operant conditioning. With the usual programmed material presented by a write-in machine or a programmed text, the subject controls the interval between S_1 -R- S_2 while the programmer has control over the interval between S_2 and S_1 . As soon as the subject has completed R and turned to S_2 the next frame S/b represents a new problem.

However, it does not appear to this writer that the distinction Zeaman makes is of paramount importance. The question that does seem

to demand an answer is: Why does immediate knowledge of results significantly improve learning and retention in some instances and why does it fail to do so in other instances?

When Skinner (1954) applied his findings in animal training to human learning he accepted response -- confirmation as reinforcement, an assumption which has been tested (Perin, 1943) and which can be accepted if we allow a further assumption that humans have been conditioned to accept being right as a satisfying state, which reduces some primary need such as avoidance of unpleasant criticism, teacher or class approval or a more distant goal of successful completion of a course with subsequent employment.

Accepting the above position, the original question still has to be answered and it is felt that this answer could be provided by an application of Bruner's cognitive theory to the learning as it takes place in programmed instruction.

Bruner conceives of learning or concept attainment in terms of categorizing and building of coding systems. He defines a coding system as a "set of contingently related, non-specific categories". It is the person's manner of grouping and relating information about his world, and it is constantly subject to change and reorganization (Bruner, 1957, p. 6). Both Bartlett's memory schemata and Piaget's work on the child's conception of nature represent naturalistic accounts of coding systems in the child. Bruner makes it clear that the coding system he describes is a hypothetical construct which is inferred from consequent events, however, it appears to be useful in explaining learning in programmed instruction.

When the individual is faced with novel stimulus he has to decide how to code or categorize. He must attend to the stimulus and deal with it. There are various stages in such decision making which Bruner describes. First there is some sort of primitive categorization, this seems to be a "silent" process that results in the perceptual isolation of the event and acceptance that the event is marked by certain spatial-temporal-qualitative characteristics.

Secondly, there begins a cue search. In cases where the object or event is not immediately perceived as a phenomenon of an established category an active cue searching occurs. Here, one is scanning the environment for data in order to find cues that permit a more precise placement of the object. Under these circumstances the organism is "open" to maximum stimulation, in a manner described below.

When a tentative categorization has occurred the cue search changes. The organism will no longer be open to stimulation but will begin to search out additional confirmatory cues to check the initial placement. During this process selective gating (screening out of information) takes place and irrelevant stimulation is reduced.

The final step in this sequence Bruner (1957) calls confirmation completion. It is marked by the termination of cue searching. Openness to further cues is drastically reduced and incongruent input is normalized or "gated out".

The critical problem that arises is the question of fit between cue and category. By a category Bruner means a rule for classing objects as equivalent. The rule specifies the following about the

instances that are to be comprised in the category.

- (a) The properties or critical attribute values required of an instance to be coded in a given class.
 - (b) The manner in which such attribute values are to be combined in making an inference from properties to category membership; whether conjunctively (e.g. a_i and b_i), relationally (e.g. a_i bears a certain relation to b_i), or disjunctively (e.g. a_i or b_i).
 - (c) The weight assigned various properties in making an inference from properties to category membership.
 - (d) The acceptance limits within which properties must fall to be criterial. That is to say, from what range of attribute values may a_i b_i k_i be drawn.
- (Bruner, 1957, p. 132.)

The above rules that govern the operation of categorization are not intended to be cautious rules. The likelihood of categorization of a sensory input will not only depend on the fit between the input and the category specification but will also depend on the accessibility of a category. By accessibility Bruner (1957) refers to the learned probabilities of occurrence of events in the person's world.

The more frequently in a given context instances of a given category occur, the greater the accessibility of the category. Operationally, this means that less stimulus input will be required for the instance or event to be categorized in terms of a frequently used category (p. 132).

In learning to categorize information input, individuals make

decisions with regard to the acquisition, retention, and utilization of information. There appears to exist some regularity in this decision making which Bruner calls strategies. These strategies insure certain forms of outcome and preclude certain others.

Bruner lists the objectives of a strategy in concept attainment as follows:

- (a) To insure that the concept will be attained after the minimum number of encounters with relevant instances.
- (b) To insure that a concept will be attained with certainty, regardless of the number of instances one must test enroute to attainment.
- (c) To minimize that amount of strain on inference and memory capacity while at the same time insuring that a concept will be attained.
- (d) To minimize the number of wrong categorizations prior to attaining a concept (p. 54).

Again it must be made clear that Bruner is not concerned with the question of consciousness of these strategies. He infers strategies from the pattern of decisions one observes in a problem-solver seeking to attain a concept. What instances does he seek to test, what hypotheses does he construct, how does he change these when he meets certain contingencies? Bruner insists on the construction of ideal strategies by the individual. Such ideal strategies can be stated in strictly logical terms.

For any given concept-attainment task, for example, there is an ideal strategy that can be constructed having the property that by following it one can attain a concept with a minimum of encounters -- but without regard to cognitive strain involved. There are other ideal strategies having the property of minimizing cognitive strain but they are often wasteful of the number of instances one must encounter enroute to solution. (Bruner, 1957, p. 55)

What is clearly needed is an ideal compromise strategy that serves both cognitive economy and rapid solution. These strategies are not fixed but can be altered with the nature of the concept being taught, and with the kinds of pressures that exist in the situation.

Bruner is not concerned with how strategies are learned.

"What" is learned, however, is not of the order of a set of simple responses. For systematic behavior of subjects attaining concepts is a highly patterned, skilled performance. If contemporary theories of learning are to deal with such performances, it is our feeling that the unit of analysis now called the "response" will have to be broadened considerably to encompass the long, contingent sequence of acts that, more properly speaking, can only be called "performance". (Bruner, 1956, p. 58)

In applying this theoretical model to the programmed learning situation the process of learning can be explained as follows. The learner is faced with a frame, it includes information which constitutes input. The learner must respond to the frame in the appropriate manner using whatever clues and cues are available. This calls for the categorization of the information included in the frame and going beyond this information by responding to it. This responding is done covertly at first and the response may or may not have to be written down depending on the mode of response called for by the programmer. Next validation of the response is required, this is given by the correct response supplied by the programmer. However, the subject may be able to gather enough cues from the frame itself that will allow him to be quite sure of having categorized the information correctly or he may find more cues in a succeeding frame that would confirm his categorization of the previous one without the aid of the external criterion supplied by the programmer. Thus validation of responses

becomes a function of informational cues within the frame plus external criterion against which responses can be checked, i.e. the answer or a later frame.

Thus reinforcement schedules in the Skinnerian sense become a function of the learner's ability to delay checking his coding against externally supplied criteria or the learner's ability to do without any formally supplied criteria, the only confirmation check being the cues in the frames themselves or the succession of frames in the order that subsequent frames are dependent on successful completion of previous frames. It may be hypothesized that in a "well-written" linear program enough cues are provided in the sequence and structure of the material that the learner can categorize and be confident of correct categorization without any answers given by the programmer. This seems especially true where the learner has already some appropriate categories in the field he is learning and where he has learned to use appropriate strategies, i.e. learning Science IX after having taken Science in the lower grades.

It may not be true of subject fields that are entirely new to the learner, that is, in which he has few categories or a limited performance repertoire and has little or no experience with strategies. It may also not explain the best method for rote and skill learning.

Skinner in his many experiments has worked with two facets of behavior organization, which have not always been clearly differentiated: (a) "shaping" of complex behavior and (b) building up rates of responding and building persistence (non reward tolerance) in response. The teaching-machine situation is an example of shaping of behavior, in

that a final complex behavior organization is approached by successive and progressive approximations. This is the principle of operant conditioning which Skinner proposed to apply to human learning. In this procedure -- in the shaping of a response -- continuous reinforcement is utilized. However, once a response has been shaped, this persistence of the overall response and increase in its rate of occurrence can be promoted by instituting a schedule of reinforcement for the overall response, i.e. by not reinforcing it every time it occurs but according to some ratio schedule.

If an analogy can be drawn between response shaping or formation in pigeons and the acquisition of a relatively complex response involving the higher mental processes in humans, then implications for the use of teaching machines are clear. The gradual acquisition of a higher mental process behavior pattern by increasing in small steps the complexity of the response should involve continuous reinforcement. Schedule of reinforcement may, however, be important in training which employs automatic devices in two respects, the first of which Skinner seems to have overlooked but would have certainly been incorporated into his theoretical framework.

(a) To maintain an optimal level of device operation in the acquisition of the shaped behavior, intermittent reinforcement should be used. The slot machine analogy will illustrate this point. For learning to occur the device will have to be operated and a very high level of operation may be maintained by fairly frequent small rewards and very infrequent large rewards. It must, however, be pointed out that machine operation itself does not insure learning, yet it may keep

the learner at the task longer, therefore, other motivational factors must also be included.

(b) The second possibility of reinforcement schedules in training is related to building up rate and persistence of the shaped behavior. Thus schedules of reinforcement would come after the response had been shaped.

The above discussion represents a stimulus-response explanation of the use of schedules of reinforcement in programmed learning. There exists a wealth of experimental data to support this theoretical framework in the case of subhuman organisms and in the learning of motor skills and simple memory training. However, it is proposed that while research in scheduled reinforcement supports this framework it apparently cannot be applied advantageously to learning that involves the higher mental processes. Bruner's explanation applies much better in the area of cognitive learning than does Skinner's reinforcement theory, learning for Bruner is explained in terms of concept attainment, outlined above.

The learner in most school subjects and in programmed instruction is engaged in formulating concepts and categorizing information. As pointed out before, schedules of reinforcement become schedules of knowledge of results which the learner uses to check the correctness of his categorization. Bruner has pointed out that human beings constantly go beyond the information given (Bruner, 1957). This applies also to programmed instruction where the learner goes beyond the information given in the frame to respond to the frame with less than certain knowledge of the correct answer. When he checks his answer against the

feedback provided by the correct response and succeeding frames he checks his categorization and finds out if he went beyond the information given in a correct form. Now the problem of scheduled reinforcement becomes important, the programmer must decide if constant or intermittent knowledge of results is more appropriate. It appears that withholding knowledge of results at certain intervals will lead the learner to become more "cue-conscious". In his search for confirmation he will attend more closely to the information given in the frame and in the succeeding frames in order to find confirmation of his categorizations. A further reason for providing knowledge of results according to a variable ratio schedule is that it has been found with human beings that knowledge of results can be delayed for a longer time than can rewards be withheld with subhuman organisms (Perin, 1943).

A further factor that has to be considered in connection with scheduled knowledge of results is the influence of previously established categories or previous knowledge in the area of study. Thus the student studying Grade X Science has been exposed to Science in his previous schooling and has acquired a set of categories which he brings with him to the new learning situation. He has therefore some readymade categories into which he can categorize the new material. Therefore, he may not need an externally supplied criteria against which to check his work at every step.

In summary, then, schedules of knowledge of results, which Skinner equates to schedules of reinforcement, may profitably be used

in programmed instruction, when the learning involves the higher mental processes and when fairly mature learners are involved. Schedules of feedback will cause the learner to be more attentive to information in the program and will increase his motivation to go through a portion of the program in order to check on previous categorizations.

Skinner's objective in using scheduled reinforcement to build up the rate and persistence of response becomes meaningless in the light of the present discussion. As Luria (1961) has pointed out once a verbal rule has been established no further reinforcement is necessary. In Brunerian terms, once a category has been established and confirmed no additional information is necessary to give strength to the category. Therefore, if we deal with the higher mental processes little is to be gained from scheduled knowledge of results or knowledge of results per se after a concept has been attained by the learner.

CHAPTER IV

THE PROCEDURE

The research into the effectiveness of reinforcement as it applies to linear small step programs has not been conclusive. Some of the studies reported in the literature and reviewed in Chapter II are plagued by methodological weaknesses, others report contradictory findings. The use of very short programs, small numbers of subjects and usually only college students in their investigation have led this writer to question some of the assertions made by the proponents of immediate continuous programs supplied knowledge of results.

The purpose of this study is to investigate the relationship of different amounts of knowledge of results to the effectiveness with which subjects learn from a linear small step program. An attempt was made to use a fairly lengthy program at a level of difficulty well below the college level and that could be integrated into the school curriculum in a meaningful manner.

METHOD AND EXPERIMENTAL DESIGN:

The program chosen for this study was A Program on Earth - Sun Relations; Developmental Edition by Robert A. Saveland and published by Ginn and Company, Toronto. The program consisted of eight sets covering the following topics.

Set 1. Earth Movements - Rotation and Revolution.

Set 2. Sun Lines - Arctic Circle and Antarctic Circle.

Set 3. Sun Lines - Tropic of Cancer and Tropic of Capricorn.

Set 4. Review of Latitude.

Set 5. Review of Longitude.

Set 6. Longitude and Time.

Set 7. Using a Time Zone Map.

Set 8. Solving Time Problems.

The publishers recommend that each set be studied in one class period. The program consisted of 213 frames written at a reading grade level of 6.1 and is recommended for Junior High School Grades. It is a good example of the Skinnerian linear type program, using very small step size and depending on a sequential progression through all frames. Each frame relies on the previous one, often using the response of a previous frame in the text of succeeding frames. The program was presented in book form with frames reading down the page and answers to the frames appearing with the next frame. A mask was supplied to cover up about 1/3 of the page with a window through which the frames and the answers could be seen. Some of the frames required the student to provide a missing word, others required some calculations. Responses were recorded in a special answer booklet which could be scored at the end of the experiment. Junior High School teachers who had examined the program recommended that it would fit best into the Grade VIII Social Studies curriculum at the beginning of the school term. A pretest and a posttest, which contained 50 items each, were also supplied by the programmer, these were parallel forms. The test questions consisted of simple computation and "fill-in" items testing comprehension and applications of material learned.

THE SAMPLE:

Four classes of Grade VIII students at the Ponoka Junior High School were involved in this experiment. This represented the total Grade VIII population and involved 125 subjects. Of these, three had to be eliminated because they missed two or more days during the time the experiment was conducted. There were 60 boys and 62 girls left in the experiment, ranging in age from 12 years, 10 months to 15 years, 6 months. Their intellectual ability ratings on the Henmon-Nelson Intelligence Test ranged from 73 to 134 with an average of 109.836. The students came from a small town and the surrounding area.

PROCEDURES:

Subjects were assigned randomly, using a table of random numbers, into one of four treatment groups: w, x, y, and z.

On the Friday prior to working with the program subjects were given the pretest after they had been told that they would participate in an experiment testing a new method in learning. On Monday the regular classroom teachers distributed the programs and answer booklets with their name and group identification, w, x, y, or z, to keep them throughout the following week and to make all of their responses in the answer booklet. They were also instructed in the use of the mask with the program, but no attempt was made to discourage them from looking at the correct response before writing it down. However, teachers observed that the students were surprisingly "honest" in responding and not checking their answers with the one provided by the programmer before recording it.

Subjects were allowed to proceed at their own rate and asked

to return the program at the end of each regular class period. The next day the programs were passed out again and subjects proceeded to carry on their work until they had completed all frames or until the end of the week, whichever came first. Only a small number (12) of subjects did not complete the program by the end of the week.

On the following Monday the posttest was administered to all subjects during the regular class period.

STATISTICAL ANALYSIS:

The statistical analysis chosen for this study follows the Multiple Linear Regression method developed by Bottenberg and Ward (1963). This technique seeks to clarify whether or not a critical variable, when added to a linear expression, significantly reduces the criterion error sum of squares.

The general approach is to express the vector of criterion variable data as a linear combination of a set of predictor vectors.

$$y = a_1x^{(1)} + a_2x^{(2)} + a_3x^{(3)} + \dots + e$$

Where y = vector of criterion variable data

$a_1a_2a_3$ = unknown weights associated with the predictor vectors

e = error residual vector.

The problem is to find a set of weights which minimize the sum of squares of the elements of vector e . These weights are referred to as "least square weights".

We wish to minimize:

$$\sum_{i=1}^N (e_i)^2 = (e_1)^2 + (e_2)^2 + (e_3)^2 + (e_4)^2 + \dots + (e_n)^2$$

To test the hypotheses in this study models such as those illustrated below were used.

Full Model

$$Y = a_0 u + a_1 x^{(1)} + a_2 x^{(2)} + a_3 x^{(3)} + a_4 x^{(4)} + e_F$$

Restricted Model

$$Y = a_0 u + a_1 x^{(1)} + a_2 x^{(2)} + e_R$$

where Y = criterion vectors

u = unit vector

$x^{(1)}$, $x^{(2)}$, $x^{(3)}$, and $x^{(4)}$ are the predictor vectors

a_0 , a_1 , a_2 , a_3 , and a_4 are the least square weights associated with the predictor vectors.

e = residual or error vector.

In testing the hypotheses the full model is compared with the restricted model. The restriction (omission of vectors $x^{(3)} + x^{(4)}$ from the full model in the restricted one) serves as a basis for calculation of whether the data in vectors $x^{(3)}$ and $x^{(4)}$ have the effect of significantly changing the value of the criterion Y in the presence of vectors $x^{(1)}$ and $x^{(2)}$.

The hypotheses are stated in the null form $\sum (e_F)^2 = \sum (e_R)^2$ which means that the error sum of squares obtained by using the full

model is equal to the sum of squares obtained from the restricted model.

The hypotheses are then tested by using the F statistic.

$$F = \frac{q_2 - q_1 / (m_1 - m_2)}{q_1 / (N - m_1)} = \frac{R_F - R_R / (m_1 - m_2)}{1 - R_F^2 / (N - m_1 - 1)}$$

where q_1 = minimized error sum of squares obtained from the full model expressing the observed values in vector Y as a linear combination of an unrestricted set of predictors

$$\left[(e_F)^2 \right]$$

where q_2 = minimized error sum obtained from the restricted model expressing the observed values in vector Y as a linear combination of a restricted set of predictors. $\left[(e_R)^2 \right]$

m_1 = number of unknown weights in full model

m_2 = number of unknown weights in restricted model

R_F^2 = Variance accounted for in full model

R_R^2 = Variance accounted for in restricted model

n = number of cases.

We are therefore testing whether or not reduction in error sum of squares obtained from the full model is significantly greater than the error sum of squares from the restricted model.

DESCRIPTION OF VECTORS EMPLOYED IN THIS STUDY

1. Sex.
2. IQ score, Henmon-Nelson Test of Intellectual Ability.
3. Pretest score - publisher supplied test.

4. Posttest score - publisher supplied test parallel to pretest in content and difficulty.
5. Treatment group membership either w, x, y, or z.

The statistical analysis developed by Bottenberg and Ward requires that categorical vectors are generated. Therefore, two vectors were generated for sex membership. Vector 6 = 1 if subject is female, 0 if otherwise; vector 7 = 1 if subject is male, 0 if otherwise. The same procedure was undertaken for treatment groups.

Vector 8 = 1 if treatment W, 0 otherwise

Vector 9 = 1 if treatment X, 0 otherwise

Vector 10 = 1 if treatment Y, 0 otherwise

Vector 11 = 1 if treatment Z, 0 otherwise

Eight further vectors were generated to test the sex treatment interaction.

Sex	F	Vector ⁽¹²⁾	Vector ⁽¹³⁾	Vector ⁽¹⁴⁾	Vector ⁽¹⁵⁾
	M	Vector ⁽¹⁶⁾	Vector ⁽¹⁷⁾	Vector ⁽¹⁸⁾	Vector ⁽¹⁹⁾

Vector 12 = 1 if subject is female and in treatment Group W,
0 otherwise

Vector 13 = 1 if subject is female and in treatment Group Y,
0 otherwise

Vector 14 = 1 if subject is female and in treatment Group Y,
0 otherwise

Vector 15 = 1 if subject is female and in treatment Group Z,
0 otherwise

Vector 16 = 1 if subject is male and in treatment Group W,
0 otherwise

Vector 17 = 1 if subject is male and in treatment Group X,
0 otherwise

Vector 18 = 1 if subject is male and in treatment Group Y,
0 otherwise

Vector 19 = 1 if subject is male and in treatment Group Z,
0 otherwise

Since the original information in vectors 1 and 5 has been generated into vectors 6, 7 and 8, 9, 10, 11 which represent categorical vectors in which membership in one means exclusion from any other, these vectors are no longer considered.

Vector 2 - IQ

3 - Pretest

4 - Posttest are continuous vectors

CHAPTER V

RESULTS

From the information collected on each of the subjects and by using the Multiple Linear Regression method developed by Bottenberg and Ward (1963), it was possible always to develop hypotheses that could be tested by the procedure outlined above. The hypotheses can be grouped under headings which appear below. For each hypothesis the two regression equations are given in the Appendix (page 79) also the vectors involved are listed.

Tests for Sampling Bias on the Control Factors

With experimental groups as small as those used in this study the random assignment of subjects to treatment groups does not ensure that the resulting groups will be equal with respect to antecedent characteristics which are presumed to be related to the criterion variable. Therefore, hypotheses I through VI were developed to test the significance of any differences between the sexes and among the treatment groups on prior knowledge of the subject matter (pretest score) and intelligence (Henmon-Nelson Test scores). Hypotheses I through III are in reference to differences on intelligence and hypotheses IV through VI are in reference to prior knowledge.

Differences Among the Groups on Intelligence Test Scores

The average intelligence test scores made by the subjects of this study classified by treatment group and sex are shown in Table 1.

Inspection of this table will reveal that girls exceeded the boys in average intelligence. There was also a difference among the treatment groups on intelligence. Treatment group W averaged 111.74 for the highest and the Z group 108.65 for the lowest intelligence score averages. Girls in treatment group W had the highest average intelligence (115.60), whereas boys in the Y treatment group had the highest average intelligence (109.33). The lowest average for girls was obtained by the girls in treatment group Z (108.285), whereas the lowest average intelligence scores for boys were found in treatment group X (106.50).

Table 1

Mean Intelligence Test Scores by Sex and Treatment

Sex	Treatment Groups				All Treatment Groups
	W	X	Y	Z	
Boys	108.125	106.5	109.333	108.941	108.33
Girls	115.6	110.612	111.126	108.285	111.419
Boys & Girls	111.741	108.966	110.233	108.645	109.901

To determine whether the differences among the treatment and sex groupings on intelligence, which were noted in Table 1, were of significance three hypotheses were developed. They were:

Hypothesis I: There are no significant differences
among the treatment groups on intelligence.

Hypothesis II: There are no significant differences

between the sexes on intelligence.

Hypothesis III: There are no significant differences in average intelligence among the sex by treatment subgroupings.

These hypotheses were tested using the procedure described on pages 47 to 50. The results of these tests are summarized in Table 2. A complete presentation of the full and restricted models for these and ensuing analyses together with the variables included in them are shown in the Appendix.

Table 2 shows that none of the F ratios for the first three hypotheses reached the necessary level of significance to reject the hypotheses. Therefore, it must be accepted that while there are differences in average intelligence scores between the sexes and the treatment groups (see Table 1) these are not large enough to render further analysis invalid.

Table 2

Results for Hypotheses I through III Summarized

Hypotheses	Criterion	Model	Vectors	R ²	F	Significance level of F
I	IQ	full	u,8-11	0.0100	0.399	N.S.
		restricted	u,99	0.000		
II	IQ	full	u,6-7	0.0197	2.409	N.S.
		restricted	u,99	0.000		
III	IQ	full	u,12-19	0.0495	0.754	N.S.
		restricted	u,6-7,8-11	0.0306		

Differences Among the Groups on Previous Knowledge of Material

Previous knowledge of material to be studied by working through the program was measured by the pretest. Scores on the pretest varied considerably from a low of 8 to a high of 40 with a mean of 24.122 out of a possible 50. Table 3 shows the average pretest scores arranged by sex and treatment groupings. Boys did slightly better than girls with a mean of 24.520 as compared to the girls with 23.268. There are also differences between the treatment groups with treatment group X having the highest average pretest score (25.333) and treatment group Z being the lowest (23.161). Furthermore average pretest scores varied within the sex groupings, boys in treatment group X score the highest average pretest score while those in treatment group Z have the lowest. The girls in treatment group X share with the boys the highest average pretest score but for girls the lowest average is found in treatment group W. Although some of the differences are very small others appear large enough to warrant careful investigation.

Table 3

Mean Pretest Score by Sex and Treatment

Sex	Treatment Groups				All Treatment Groups
	W	X	Y	Z	
Boys	24.812	26.083	25.600	23.588	24.520
Girls	22.400	24.833	23.200	22.642	23.268
Boys & Girls	23.645	25.333	24.400	23.161	24.122

To determine whether the differences among treatment and sex groupings on pretest scores which were noticed in Table 3 (above) were

significant, three hypotheses were developed. They were:

Hypothesis IV: There are no significant differences among treatment groups on pretest performance.

Hypothesis V: There are no significant differences between the sexes on pretest performance.

Hypothesis VI: There are no significant differences in average pretest performance among the sex by treatment subgroupings.

The resulting analysis (see Appendix) fails to show any significant differences between the sexes, among the treatment groups or among the sex by treatment subgroupings. The results of the tests for the hypotheses IV through VI are found in Table 4 in summarized form. Because none of the F ratios reached significance, all three hypotheses had to be accepted. This ruled out any biasing influence with reference to previous knowledge or familiarity with the material to be studied and allowed further investigation of the results of this study.

Table 4

Results for Hypotheses IV through VI Summarized
=====

Hypo- theses	Cri- terion	Model	Vectors	R ²	F	Signifi- cance level of F
IV	Pretest	full	u,8-11	0.0090	0.3585	N.S.
		restricted	u,99	0.000		
V	Pretest	full	u,6-7	0.0224	2.7513	N.S.
		restricted	u,99	0.000		
VI	Pretest	full	u,12-19	0.0370	0.0977	N.S.
		restricted	u,6-7,8-11	0.0345		

Differences Among the Groups on Posttest Scores

Hypotheses VII through IX are designed to test the effect the treatment (working through the program with differences in schedules of knowledge of results) had on posttest performance. The posttest is a parallel form of the pretest (see Chapter IV, p. 48). Table 5 shows the average posttest score classified by sex and treatment groups. Inspection of Table 5 reveals that the average posttest scores for boys and girls is very similar, the difference is only 0.05 points. However, there are differences among the treatment groups with treatment group W achieving the highest average posttest score and treatment group Z being the lowest (33.548), with treatment group X also very close to the scores of group Z (33.666). Within the sex by treatment subgroupings it will be noted that while boys in the Y treatment group earned the highest average posttest score, girls in the same treatment group earned the lowest average posttest score. Boys in treatment group X had the lowest average posttest scores (32.916) while the girls in the same treatment group score the second highest average for their groups (34.166). The girls in treatment group W have the highest average score (35.666) which is identical to the highest average for the boys which was found in treatment group Y.

To determine whether the differences among the treatment and sex groupings on posttest performance which are noted in Table 5 (see the following page) were significant, three hypotheses were developed. They were:

Hypothesis VII: There are no significant differences among the treatment groups on posttest performance.

Hypothesis VIII: There is no significant difference between the two sexes on posttest performance.

Hypothesis IX: There are no significant differences in average posttest score among the sex by treatment subgroups.

Table 5

Mean Posttest Scores by Sex and Treatment

Sex	Treatment Groups				All Treatment Groups
	W	X	Y	Z	
Boys	34.687	32.916	35.666	33.058	35.116
Girls	35.666	34.166	32.200	34.142	34.066
Boys & Girls	35.161	33.666	33.933	33.548	34.081

The hypotheses were tested and the results of these tests are summarized in Table 6. For all three hypotheses the results were not significant and the above hypotheses can be accepted. Hypothesis VII, referring to the effect of the treatment, and Hypothesis IX, in which treatment also appears, are key hypotheses in this study.

Table 6

Summary of Tests of Hypotheses VII through IX

Hypotheses	Criterion	Model	Vectors	R ²	F	Significance level of F
VII	Posttest	full restricted	u,8-11 u,99	0.0074 0.000	0.2922	NS
VIII	Posttest	full restricted	u,6-7 u,99	0.0000 0.000	0.0000	NS
IX	Posttest	full restricted	u,12,17-19 u,6-7, 8-11	0.0225 0.0074	0.5867	NS

Posttest Performance with Intelligence Controlled

Previous analyses have shown that there were no significant differences on intelligence among the various treatment groups, between the sex groups or among the sex by treatment subgroups. However, it is recognized that performance on a posttest of achievement is related to intelligence. Thus, some portion of the variation among individuals within groups on posttest performance can be accounted for by the variation among them on intelligence. A test of significance such as the one used herein is based upon the ratio of between-groups variation to within group variation. It is possible that if within-group variations in posttest performance which are associated with within-group variations on intelligence are partialled out through control on intelligence, the denominator of the ratio will be reduced. The previous analyses showed no significant among-groups variations on intelligence. Therefore, it is possible that

the resulting ratio will be enlarged by partialling out the within-group variability. Consequently, control on intelligence will allow the test of significance to be sharpened. Three hypotheses were developed in which intelligence has been controlled. They were:

Hypothesis X: There are no significant differences among treatment groups on posttest performance when differences among them on intelligence have been controlled.

Hypothesis XI: There is no significant difference between the two sexes on posttest performance when differences between them on intelligence have been controlled.

Hypothesis XII: There are no significant differences on average posttest performance among the sex by treatment subgroups when differences among them on average intelligence scores have been controlled.

The tests for the above hypotheses are shown in full in the Appendix and are summarized in Table 7 below.

Table 7

Summary of Tests of Hypotheses X through XII

=====

Hypo- theses	Cri- terion	Con- trolled	Model	Vectors	R ²	F	Signifi- cance level of F
X	Posttest	IQ	full restricted	u,2,8-11 u,2	0.402 0.401	0.0762	NS

cont'd...

XI	Posttest	IQ	full	u,2,6,7	0.409	1.6319	NS
			restricted	u,2	0.401		
XII	Posttest	IQ	full	u,2,12-19	0.427	1.6821	NS
			restricted	u,2	0.401		

Inspection of Table 7 reveals that the three hypotheses cannot be rejected. Therefore, when the effect of intelligence has been partialled out there are no significant differences on residual posttest performance among the treatment groups, between the two sexes and among the sex by treatment subgroups.

Posttest Performance with Pretest Controlled

In the three preceding hypotheses control on intelligence was exercised; however, previous knowledge of the material would also have an effect on posttest performance. Hypotheses IV through VI (page 59) showed no significant differences among the treatment groups, between the two sexes and the sex by treatment subgroups on average pretest scores.

However, it is possible that some portion of the variation among individuals within groups on posttest performance can be accounted for by the variation among them on pretest score (previous familiarity with the subject matter). It is possible that if within-group variations in posttest performance which are associated with within-group variations on pretest are partialled out through control on pretest the denominator of the ratio will be reduced. The previous analyses showed no significant among-groups variation on pretest. Consequently, control on pretest will allow the test of significance to be sharpened.

The following three hypotheses were developed to partial out the residual effect of pretest performance on posttest. They were:

Hypothesis XIII: There are no significant differences among the treatment groups on posttest performance when differences among them on pretest have been controlled.

Hypothesis XIV: There is no significant difference between the two sexes on posttest performance when differences between them on pretest have been controlled.

Hypothesis XV: There are no significant differences in average posttest scores among the sex by treatment subgroups when the differences among them on average pretest scores have been controlled.

The tests for the above hypotheses have been summarized in Table 8.

Table 8

Summary of Tests of Hypotheses XIII through XV

Hypo- theses	Cri- terion	Con- trolled	Model	Vectors	R^2	F	Signifi- cance level of F
XIII	Posttest	Pretest	full restricted	u,3,8-11 u,3	0.547 0.533	1.2434	NS
XIV	Posttest	Pretest	full restricted	u,3,6-7 u,3	0.545 0.533	3.1819	NS
XV	Posttest	Pretest	full restricted	u,3,12-19 u,3	0.576 0.533	3.7991	0.01225

On the basis of the results of the tests for these hypotheses, hypotheses XIII and XIV cannot be rejected. However, hypothesis XV must be rejected. There does exist a significant difference among the sex by treatment subgroups on residual posttest performance when the effect of previous knowledge (pretest) has been partialled out. In order to determine the differences between the sex by treatment subgroups expected average posttest scores for each of the subgroups were compared with actual average posttest scores. The results appear in Table 9.

Table 9

Actual and Expected Mean Posttest Scores and the Differences Between Them by Sex and Treatment Group with Control on Intelligence

=====

	W			X		
	A	P	D	A	P	D
Boys	34.687	34.928	-.241	32.916	36.131	-3.215
Girls	35.666	32.646	+3.020	34.166	34.948	-.782
	Y			Z		
	A	P	D	A	P	D
Boys	35.666	35.674	-.008	33.058	33.770	-.712
Girls	32.200	33.403	-1.203	34.142	32.876	+1.266

Inspection of Table 9 reveals that treatment W was best for girls while treatment Y was best for boys. Girls scored 3.020 higher than expected, boys scored just as expected. Treatment X proved least suited for boys, they scored 3.215 below the expected,

while treatment Y appears to be least suited for the girls, they scored 1.203 below the expected average posttest score.

Posttest Performance with Control on Both Intelligence and Pretest Performance

In the final three hypotheses the residual effect of both intelligence and previous knowledge are partialled out. This procedure would tend to sharpen the tests for the effect of treatment and would constitute the most stringent test of it because the residual effects of intelligence and previous knowledge would be controlled. The following hypotheses were developed. They were:

Hypothesis XVI: There are no significant differences among treatment groups on posttest performance when differences among them on intelligence and pretest have been controlled.

Hypothesis XVII: There is no significant difference between the two sexes on posttest performance when differences between them on intelligence and pretest have been controlled.

Hypothesis XVIII: There are no significant differences in average posttest scores among the sex

by treatment subgroups when differences among them on average intelligence and pretest scores have been controlled.

The tests for the above hypotheses are summarized in Table 10.

Table 10

Summary of Tests of Hypotheses XVI through XVIII

Hypo- theses	Cri- terion	Con- trolled	Model	Vectors	R ²	F	Signifi- cance level of F
XVI	Posttest	IQ	full	u,2,3,8-11	0.630	0.7811	NS
		Pretest	restricted	u,2,3	0.623		
XVII	Posttest	IQ	full	u,2-3,6-7	0.624	0.3723	NS
		Pretest	restricted	u,2,3	0.623		
XVIII	Posttest	IQ	full	u,2,3,12-19	0.643	2.0931	NS
		Pretest	restricted	u,2,3	0.623		

Inspection of Table 10 reveals that none of the three hypotheses can be rejected. When the residual effects of intelligence and previous knowledge have been partialled out there are no significant differences on posttest performance among the treatment groups, between the two sexes and among the sex by treatment interaction subgroups.

The hypotheses are summarized in the following section.

Summary of Results of Analysis

Hypothesis Number	Hypothesis Statement	Result of Analysis
I	There are no significant differences among the treatment groups on intelligence	accepted

II	There are no significant differences between the sexes on intelligence	accepted
III	There are no significant differences in average intelligence among the sex by treatment subgroupings	accepted
IV	There are no significant differences among treatment groups on pretest performance	accepted
V	There are no significant differences between the sexes on pretest performance	accepted
VI	There are no significant differences in average pretest performance among the sex by treatment subgroupings	accepted
VII	There are no significant differences among the treatment groups on posttest performance	accepted
VIII	There are no significant differences between the sexes on posttest performance	accepted
IX	There are no significant differences in average posttest performance among the sex by treatment subgroupings	accepted
X	There are no significant differences among treatment groups on posttest performance when differences among them on intelligence have been controlled	accepted
XI	There are no significant differences between the sexes on posttest performance when differences between them on intelligence have been controlled	accepted

- | | | |
|------|---|----------|
| XII | There are no significant differences on average posttest performance among the sex by treatment subgroups when differences among them on average intelligence scores have been controlled | accepted |
| XIII | There are no significant differences among the treatment groups on posttest performance when differences among them on pretests have been controlled | accepted |
| XIV | There are no significant differences between the two sexes on posttest performance when differences between them on pretest have been controlled | accepted |
| XV | There are no significant differences in average posttest score among the sex by treatment subgroups when the differences among them on average pretest score have been controlled | rejected |
| XVI | There are no significant differences among treatment groups on posttest performance when differences among them on intelligence and pretest have been controlled | accepted |
| XVII | There is no significant difference between the two sexes on posttest performance when differences between them on intelligence and pretest have been controlled | accepted |

XVIII There are no significant differences in accepted
average posttest scores among the sex by
treatment subgroups when differences among
them on average intelligence and average
pretest scores have been controlled

DISCUSSION

As the result of the analysis of the data and the testing of the hypotheses it has to be accepted that knowledge of results, as provided by the programer, had no effect on the subjects' ability to learn from this program, A Program on Earth-Sun Relations.

Hypotheses I through VI tested for possible biasing effects of intelligence and previous knowledge or familiarity with the material. It was found none existed. The hypotheses VII through IX tested the effect of treatment which consisted of 100%, 50%, 25%, 12½% of K of R dispensed at a variable ratio. The hypotheses had to be accepted in their null form for the difference in treatments and the two sexes.

Further hypotheses were developed to test the effect of treatment with control on intelligence, pretest and the combined effects of intelligence and pretest. With the exception of one hypothesis all had to be accepted. Hypothesis XV had to be rejected. It involved the effect of treatment by sex membership when the effect of previous knowledge had been controlled. This would indicate that intelligence in the absence of the effect of previous knowledge

tends to influence performance on the posttest.

However, when both pretest and intelligence were controlled the effect of treatment was not significant. Boys and girls performed equally well under all four treatment conditions. These results are in agreement with the findings in current literature. The studies that have been involving the higher mental processes and using similar procedures such as Moore and Smith (1961), Krumboltz and Weisman (1962), and Evans, Glaser and Homme (1962) have reported similar findings. Those studies involving simpler tasks such as bar pressing and maze learning and carried out on lower species such as rats and pigeons report greater dependence on reinforcement and knowledge of results. The results of studies by Perin (1943) involving rats establish quite clearly that reinforcement has to be provided without delay to affect learning. Similar results were obtained by Terrell and Warren (1961) using children in four discrimination tasks. Saltzman (1951) using a verbal maze as the task found that when knowledge of results was delayed a greater number of trials were necessary for learning than if knowledge of results were not delayed.

Saltzman also supports the assumption, which underlies this study, that knowledge of the correctness of the response is reinforcing for human subjects. The same assumption has been made by other investigators into the effect of knowledge of results and is basic to the Skinner theory of program writing.

Other studies discussed in the review of the research using programmed material as the task factor such as Meyer (1960) and

Teldhusen and Brit (1962) do not contradict the findings in this study but had to be rated inconclusive because of methodological weaknesses.

The results obtained support the theoretical framework as outlined in Chapter III. As long as the learner finds enough cues to confirm his categorizations and as long as he can continue this process learning will take place. In the program used in this study the learners could maintain adequate learning independently from the amount of knowledge of results, as provided by the program. Knowledge of results, however, was not absent but could be obtained from the structure of the program and through the dependence of frames on previous frames. This difficulty in control of informational feedback and intrinsic knowledge of results is a common feature of all studies using a linear small step program but has not been noted by others --Moore and Smith (1961) and Krumboltz and Weisman (1962).

One aspect of the theory advanced in this study failed to be supported. Subjects did not perform better as a result of fewer instances of knowledge of results. They may have had to be more cue conscious and more alert in order to confirm their categorizations.

BIBLIOGRAPHY

- Alter, Millicent and Silverman, Robert L., "The response in Programmed Instruction," The Journal of Programmed Instruction, 1962, 1.
- Amsel, Abram, "Error Responses and Reinforcement Schedules in Self-Instructional Devices," Teaching Machines and Programmed Learning -- a source book, Lumsdaine, A.A., and Glaser, Robert (Ed.) Washington, D.C., 1960, Department of Audio-Visual Instruction, National Education Association.
- Angell, C.W. and Troyer, M.E., "A new self-scoring device for improving instruction," School and Society, 1948, 67, 84-85.
- Angell, C.W., "The effect of immediate knowledge of quiz results on final examination scores in freshman chemistry," Journal of Educational Research, 1948, 42, 391-394.
- Auble, D. and Mech, E.V., "Quantitative studies of verbal reinforcement in classroom situations: Differential reinforcement related to the frequency of error and correct responses," Journal of Psychology, 1953, 35, 307-312.
- Auble, D. and Mech, E.V. "Partial verbal reinforcement related to distributed practice in a classroom situation," Journal of Psychology, 1953, 36, 165-186.
- Berg, I.A., "The use of human subjects in psychological research," American Psychologist, 1954, 9, 108-111.
- Bottenberg, Robert A., Ward, J.H., Jr., Applied Multiple Linear Regression, 6570th Personnel Research Laboratory, Aerospace Medical Division, Air Force System Command, Lackland Air Force Base, Texas.
- Bruner, J.S., Goodnow, Jacqueline J., Austin, George A., A Study of Thinking, New York: John Wiley and Sons, Inc., 1956.
- Bruner, Jerome S., "Going beyond the information given," Contemporary Approaches to Cognition, A Symposium Held at the University of Colorado; Harvard University Press, Cambridge, 1957(a).
- Bruner, J.S., "On Perceptual Readiness," Psychological Review, Vol. 64, No. 2, March, 1957, 123-152(b).
- Bruner, J.S., Contemporary Approaches to Cognition, "A Symposium Held at the University of Colorado," Harvard University Press, Cambridge, 1957(c).

- Bruner, J.S., The Process of Education, Harvard University Press, Cambridge, 1962.
- Bryan, G.L., Rigney, J.W., and Van Horn, C., "An evaluation of three types of information for supplementing knowledge of results in a training technique." Technical Report No. 19, prepared for the Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, at the Department of Psychology, University of Southern California, Los Angeles, April 1957.
- Elley, W.B. Why do programs teach? The ATA Magazine, 1963, 44, No. 4.
- Evans, J.L., Homme, L.E., and Glaser, R. The development and use of a "standard" program for investigating programmed verbal learning. Amer. Psychologist, 1960, 15, 424 (Abstract).
- Feldhusen, John F. and Birt, A. "A Study of Nine Methods of Presentation of Programed Learning Material", The Journal of Educational Research, 55: 461-466, June-July 1962.
- Fusco, G.C. Programed self-instruction: possibilities and limitations. In W.I. Smith and J.W. Moore (Eds.), Programmed Learning: Theory and Research. Princeton: D. Van Nostrand, 1962, Pp. 227-230.
- Harvey, O.J., Hunt, David E., Schroder, Harold M., Conceptual Systems and Personality Organization, New York: John Wiley and Sons, Inc., 1961.
- Hebb, D.O., "Elementary School Methods," Teacher's Magazine, Montreal: 1930, 12, 23-32.
- Hebb, D.O., "Drives and the C.N.S.," The Psychological Review, 62, 4, July, 1955, 243-254.
- Heron, W.T. "The effects of a differential rate of reinforcement of responses to two levers," Journal of Comparative Psychology, 1942, 33, 87-96.
- Holland, James G., "Teaching Machines: An Application of Principles from the Laboratory," Programmed Learning, Smith, Wendell J. and Moore, William J. (Ed.), Programmed Learning, D. Van Nostrand Company, Inc., (Canada) Ltd., 1962.
- Hull, C.L., Principles of Behavior, New York: Appleton Century Crofts, 1943.

- Humphreys, L.G., "The effect of random alteration of reinforcement on the acquisition and extinction of conditioned eyelid reactions," Journal of Experimental Psychology, 1939, 25, 141-158.
- Humphreys, L.G., "Extinction of conditioned psychogalvanic responses following two conditions of reinforcement," Journal of Experimental Psychology, 1940, 27, 71-75.
- Lewis, Donald J. and Harder, Donald A., "Partial Reinforcement following a long continuous reinforcement acquisition series," Journal of Genetic Psychology, 1959, 95, 207-210.
- Little, J.K., "Results of use of machines for testing and for drill, upon learning in educational psychology," Journal of Experimental Education, 1934, 3, 45-49.
- Logan, F.A., "The role of delay of reinforcement in determining reaction potential," Journal of Experimental Psychology, 1952, 43, 393-399.
- Lumsdaine, A.A. and Robert Glaser (eds.) Teaching Machines and Programed Learning: A Source Book. Washington, D.C.: National Education Association, 1961.
- Luria, Aleksander Romanovich, The Role of Speech in the Regulation of Normal and Abnormal Behavior, Edited by J. Tizard, New York: 1961, Pergaman Press.
- McGlock, J.A., The Psychology of Human Learning, New York: Longmans, 1942, chapter 4.
- McNeil, John D., "Programmed Instruction as a research tool in reading: an annotated case," The Journal of Programmed Instruction, 1962, 1.
- Meyer, Susan R. "Report on the Initial Test of a Junior High School Vocabulary Program" in Teaching Machines and Programed Learning, (eds.) Lumsdaine and Glaser. Washington, D.C.: Department of Audiovisual Instruction, Nat. Education Association, 1960, p. 229-46.
- Moore, William and Smith, Wendell J., "Knowledge of Results in Self-Teaching Spelling," Psychological Reports, 1961, 9, 717-726.
- Moore, William J. and Smith, Wendell J., "A comparison of several types of 'immediate reinforcement', Programmed Learning, Toronto: 1962, D. Van Nostrand Company, Inc.
- Munn, N.L., "Learning in Children," in L. Carmichael (Ed.) Manual of Child Psychology (2nd edition), New York: Wiley, 1954, 374-458.

- North, Joseph, "The Relationship of Broken Homes to the Performance of School Children," Unpublished Master Thesis, University of Alberta, Edmonton, Alberta, 1965.
- Perin, C.T., "Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction," Journal of Experimental Psychology, 1942, 30, 93-113.
- Perin, Theodore, "A quantitative investigation of the delay of reinforcement gradient," Journal of Experimental Psychology, Vol. 32, 1943, 37-51.
- Porter, D., "A critical review of a portion of the literature on teaching devices," Harvard Educational Review, 1957, 27, 126-147.
- Pressey, S.L., "A Simple device which gives tests and scores and teaches," School and Society, 1926, 23, 373-376.
- Pressey, S.L., "Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self instruction." Journal of Psychology, 1950, 29, 417-447.
- Saltzman, J.J., "Delay of Reward and Human Verbal Learning," Journal of Experimental Psychology, June 1951, 41: 437-39.
- Silverman, R.E. and Alter, M., "Note on the response in teaching machine programs," Psychological Reports, 1960, 7, 496.
- Schramm, W., Programed Instruction Today and Tomorrow, The Fund for the Advancement of Education, November 1962.
- Schramm, W., The Research on Programed Instruction, An Annotated Bibliography. U.S. Department of Health, Education and Welfare, Washington, 1964.
- Skinner, B.F., The Behavior of Organisms, New York: Appleton Century, 1938.
- Skinner, B.F., Science and Human Behavior, New York: MacMillan, 1953.
- Skinner, B.F., "Some contributions of an experimental analysis of behavior to psychology as a whole," American Psychologist, 1953, 8, 69-78.
- Skinner, B.F., "Science of learning and the art of teaching," Harvard Educational Review, 1954, 24, No. 2, 89-97.

- Skinner, B.F., "Reinforcement Today," American Psychologist, March 1958, 13, 94-99.
- Spence, K.W., "The role of secondary reinforcement in delayed reward learning," Psychological Review, 1947, 54, 1-8.
- Stolurrow, L.M., and Walker, C.C. A comparison of overt and covert response in programmed learning. J. Educ. Res. 1962, 55, 421-29.
- Strembitsky, J., An appraisal of research in programmed instruction. Master Thesis, University of Alberta, Edmonton, 1964.
- Taylor, Janet A. and Spence, K.W., "The Relationship of Anxiety Level to Performance in Serial Learning," Journal of Experimental Psychology, 1952, 44, 61-64.
- Terrell, G. and Ware, R., "Role of delay of reward in speed of size and form discrimination in childhood," Child Development, 1961, 32, 409-415.
- Warren, A.B. and Brown, R.H., "Conditioned operant response phenomena in children," Journal of Genetic Psychology, 1943, 28, 181-207.
- Williams, S.B., "Resistance to extinction as a function of the number of reinforcements," Journal of Experimental Psychology, 1938, 23, 506-522.
- Wolfe, J.B., "The effect of delayed reward upon learning in the white rat," Journal of Comparative and Physiological Psychology, 1934, 17, 1-21.
- Woodworth, R.S. and Schlosberg, H., Experimental Psychology (revised), New York: Holt, 1954.
- Yerkes, R.M. and Dodson, J.D., "The Relation of Strength to Rapidity of Habit-formation," Journal of Comparative Neurology and Psychology, 18: 459-482, 1908.

APPENDIX

Each of the eighteen hypotheses was tested by the method outlined under Statistical Analysis (page 50). Below each of the hypotheses is listed with the Full and Restricted model, the weighted vectors and the description of the vector involved.

Hypothesis I: There are no significant differences among the treatment groups on intelligence.

The regression equations for this hypothesis were:

Full Model

$$Y = 108.638u + 2.906 x^{(8)} + 0.329 x^{(9)} + 1.54 x^{(10)} + 0.000 x^{(11)}$$

Restricted Model

$$Y = 108.638u$$

Where Y = the intelligence quotient

u = unit vector

$x^{(8)}$ = 1 if treatment W, 0 otherwise

$x^{(9)}$ = 1 if treatment X, 0 otherwise

$x^{(10)}$ = 1 if treatment Y, 0 otherwise

$x^{(11)}$ = 1 if treatment Z, 0 otherwise

For the hypothesis I, the following were noted:

$$R_F^2 = 0.0100 \quad R_R^2 = 0.000 \quad F = 0.399 \text{ N.S.}$$

In this case the restricted model contains only the unit vector u, $(R_R)^2$ is arbitrarily defined as 0 and the expression for

F in this case would be

$$F = \frac{(R_F)^2 / (m_1 - m_2)}{(1 - R_F^2 / n - m_1 - 1)} \quad *$$

* See Bottenberg and Ward, p. 126

Hypothesis II: There is no significant difference between the sexes on intelligence.

The regression equations for this hypothesis were:

Full Model

$$Y = 108.1999u + 0.140 x^{(6)} + 0.000 x^{(7)}$$

Restricted Model

$$Y = 108.1999u$$

Where

Y = intelligence

u = unit vector

$x^{(6)}$ = 1 if female, 0 otherwise

$x^{(7)}$ = 1 if male, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.0197 \quad R_R^2 = 0.000 \quad F = 2.409 \text{ N.S.}$$

Hypothesis III: There are no significant differences in average intelligence among the sex by treatment subgroups.

The regression equations for this hypothesis were:

Full Model

$$Y = 109.039u + 6.544 x^{(12)} + 1.593 x^{(13)} + 2.097 x^{(14)} - 0.832 x^{(15)} - 1.326 x^{(16)} - 2.551 x^{(17)} + 0.348 x^{(18)} - 0.156 x^{(19)}$$

Restricted Model

$$Y = 106.980u + 3.337 x^{(6)} + 0.000 x^{(7)} + 2.873 x^{(8)} + 0.000 x^{(9)} + 1.617 x^{(10)} + 0.125 x^{(11)}$$

Where

Y = intelligence

u = unit vector

$x^{(12)}$ = 1 if subject female and in treatment group W,
0 otherwise

$x^{(13)}$ = 1 if subject female and in treatment group X,
0 otherwise

$x^{(14)}$ = 1 if subject female and in treatment group Y,
0 otherwise

$x^{(15)}$ = 1 if subject female and in treatment group Z,
0 otherwise

$x^{(16)}$ = 1 if subject male and in treatment group W,
0 otherwise

$x^{(17)}$ = 1 if subject male and in treatment group X,
0 otherwise

$x^{(18)}$ = 1 if subject male and in treatment group Y,
0 otherwise

$x^{(19)}$ = 1 if subject male and in treatment group Z,
0 otherwise

$x^{(6)}$ = 1 if subject female and 0 otherwise

$x^{(7)}$ = 1 if subject male and 0 otherwise

$x^{(8)}$ = 1 if treatment group W, 0 otherwise

$x^{(9)}$ = 1 if treatment group X, 0 otherwise

$x^{(10)}$ = 1 if treatment group Y, 0 otherwise

$x^{(11)}$ = 1 if treatment group Z, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.0495 \quad R_R^2 = 0.0306 \quad F = 0.754 \text{ N.S.}$$

Hypothesis IV: There are no significant differences

among treatment groups on pretest performance.

The regression equations for this hypothesis were:

Full Model

$$Y = 24.411u - 0.796 x^{(8)} + 0.256 x^{(9)} + 0.000 x^{(10)} \\ - 1.230 x^{(11)}$$

Restricted Model

$$Y = 24.411u$$

Where

Y = pretest score

u = 1 unit vector

 $x^{(8)}$ = 1 if treatment W $x^{(9)}$ = 1 if treatment X $x^{(10)}$ = 1 if treatment Y $x^{(11)}$ = 1 if treatment Z

For the hypothesis the following were noted:

$$R_F^2 = 0.0090 \quad R_R^2 = 0.000 \quad F = 0.3585 \text{ N.S.}$$

Hypothesis V: There are no significant differences

between the sexes on pretest performance.

The regression equations for this hypothesis were:

Full Model

$$Y = 24.917u - 1.884 x^{(6)} + 0.000 x^{(7)}$$

Restricted Model

$$Y = 24.917u$$

Where

Y = pretest score

u = unit vector

 $x^{(6)}$ = 1 if female, 0 otherwise $x^{(7)}$ = 1 if male, 0 otherwise

For hypothesis V the following were noted:

$$R_F^2 = 0.0224 \quad R_R^2 = 0.000 \quad F = 2.7513 \text{ N.S.}$$

Hypothesis VI: There are no significant differences in average pretest performance among the sex by treatment subgroupings.

The regression equations for this hypothesis were:

Full Model

$$\begin{aligned} Y = & 23.349u - 0.924 x^{(12)} + 0.348 x^{(13)} - 0.170 x^{(14)} \\ & - 0.704 x^{(15)} + 1.463 x^{(16)} + 2.75 x^{(17)} \\ & + 2.248 x^{(18)} + 9.250 x^{(19)} \end{aligned}$$

Restricted Model

$$\begin{aligned} Y = & 25.41u - 2.023 x^{(6)} + 0.00 x^{(7)} - 0.799 x^{(8)} \\ & + 0.476 x^{(9)} + 0.000 x^{(10)} - 1.325 x^{(11)} \end{aligned}$$

Where Y = pretest score

u = unit vector

$x^{(12)}$ = 1 if female in treatment group W, 0 otherwise

$x^{(13)}$ = 1 if female in treatment group X, 0 otherwise

$x^{(14)}$ = 1 if female in treatment group Y, 0 otherwise

$x^{(15)}$ = 1 if female in treatment group Z, 0 otherwise

$x^{(16)}$ = 1 if male in treatment group W, 0 otherwise

$x^{(17)}$ = 1 if male in treatment group X, 0 otherwise

$x^{(18)}$ = 1 if male and in treatment group Y, 0 otherwise

$x^{(19)}$ = 1 if male in treatment group Z, 0 otherwise

$x^{(6)}$ = 1 if female, 0 otherwise

$x^{(7)}$ = 1 if male, 0 otherwise

$x^{(8)}$ = 1 if treatment group W, 0 otherwise

$x^{(9)} = 1$ if treatment group X, 0 otherwise

$x^{(10)} = 1$ if treatment group Y, 0 otherwise

$x^{(11)} = 1$ if treatment group Z, 0 otherwise

For the hypothesis the following were noted:

$$R_F^2 = 0.0370 \quad R_R^2 = 0.346 \quad F = 0.0977 \text{ N.S.}$$

Hypothesis VII: There is no significant difference among the treatment groups on the posttest performance.

The regression equations for this hypothesis were:

Full Model:

$$Y = 33.570u + 1.725 x^{(8)} + 0.096 x^{(9)} + 0.335 x^{(10)} + 0.000 x^{(11)}$$

Restricted Model:

$$Y = 33.570u$$

Where Y = the criterion, posttest performance

u = unit vector

$x^{(8)} = 1$ if treatment W, 0 otherwise

$x^{(9)} = 1$ if treatment X, 0 otherwise

$x^{(10)} = 1$ if treatment Y, 0 otherwise

$x^{(11)} = 1$ if treatment Z, 0 otherwise

For the hypothesis the following were noted:

$$R_F^2 = 0.0074 \quad R_R^2 = 0.000 \quad F = 0.2922 \text{ N.S.}$$

Hypothesis VIII: There is no significant difference between the two sexes on posttest performance.

The regression equations for this hypothesis were:

Full Model

$$Y = 34.115u + 0.000 x^{(6)} + 0.000 x^{(7)}$$

Restricted Model

$$Y = 34.115u$$

Where Y = the criterion, posttest performance

u = unit vector

$x^{(6)}$ = 1 if subject is female

$x^{(7)}$ = 0 if subject is male

For the hypothesis the following were noted:

$$R_F^2 = 0.000 \quad R_R^2 = 0.000 \quad F = 0.000 \text{ N.S.}$$

Hypothesis IX: There are no significant differences in average posttest score among the sex by treatment subgroupings.

The regression equations for this hypothesis were:

Full Model

$$\begin{aligned} Y = & 34.124u + 1.805 x^{(12)} + 0.000 x^{(13)} - 1.933 x^{(14)} \\ & + 0.000 x^{(15)} + 0.602 x^{(16)} - 1.207 x^{(17)} + \\ & 1.547 x^{(18)} - 1.030 x^{(19)} \end{aligned}$$

Restricted Model

$$\begin{aligned} Y = & 33.570u + 0.000 x^{(6)} + 0.000 x^{(7)} + 1.725 x^{(8)} \\ & + 0.096 x^{(9)} + 0.335 x^{(10)} + 0.000 x^{(11)} \end{aligned}$$

Where Y = the criterion, posttest performance

u = unit vector

$x^{(6)}$ = 1 if female, 0 otherwise

$x^{(7)}$ = 1 if male, 0 otherwise

- $x^{(8)} = 1$ if treatment group W, 0 otherwise
 $x^{(9)} = 1$ if treatment group X, 0 otherwise
 $x^{(10)} = 1$ if treatment group Y, 0 otherwise
 $x^{(11)} = 1$ if treatment group Z, 0 otherwise
 $x^{(12)} = 1$ if female in treatment group W, 0 otherwise
 $x^{(13)} = 1$ if female in treatment group X, 0 otherwise
 $x^{(14)} = 1$ if female in treatment group Y, 0 otherwise
 $x^{(15)} = 1$ if female in treatment group Z, 0 otherwise
 $x^{(16)} = 1$ if male in treatment group W, 0 otherwise
 $x^{(17)} = 1$ if male in treatment group X, 0 otherwise
 $x^{(18)} = 1$ if male in treatment group Y, 0 otherwise
 $x^{(19)} = 1$ if male in treatment group Z, 0 otherwise

For the hypothesis the following were noted:

$$R_F^2 = 0.0225 \quad R_R^2 = 0.0074 \quad F = 0.5867 \text{ NS}$$

Hypothesis X: There are no significant differences among treatment groups on posttest performance when differences among them on intelligence have been controlled.

The regression equations expressing the hypothesis were:

Full Model

$$Y = -15.238u + 0.449 x^{(2)} + 0.471 x^{(8)} + 0.000 x^{(9)} - 0.286 x^{(10)} + 0.000 x^{(11)}$$

Restricted Model

$$Y = -15.330u + 0.450 x^{(12)}$$

Where Y = posttest score

u = unit vector

$x^{(2)}$ = intelligence

$x^{(8)}$ = 1 if treatment group W, 0 otherwise

$x^{(9)}$ = 1 if treatment group X, 0 otherwise

$x^{(10)}$ = 1 if treatment group Y, 0 otherwise

$x^{(11)}$ = 1 if treatment group Z, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.402 \quad R_R^2 = 0.401 \quad F = 0.0762 \text{ NS}$$

Hypothesis XI: There is no significant difference between the two sexes on posttest performance when differences between them on intelligence have been controlled.

The regression equations expressing this hypothesis were:

Full Model

$$Y = -15.572u + 0.459 x^{(2)} - 1.482 x^{(6)} + 0.000 x^{(7)}$$

Restricted Model

$$Y = -15.330u + 0.459 x^{(2)}$$

When

Y = posttest performance

u = unit vector

$x^{(2)}$ = intelligence

$x^{(6)}$ = 1 if subject is female, 0 otherwise

$x^{(7)}$ = 1 if subject is male, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.409 \quad R_R^2 = .401 \quad F = 1.6319 \text{ NS}$$

Hypothesis XII: There are no significant differences on

average posttest performance among the sex
by treatment subgroups when differences among
them on average intelligence scores have been
controlled.

The regression equations expressing this hypothesis were:

Full Model

$$Y = -16.834u + 0.463 x^{(2)} - 0.772 x^{(12)} - 0.270 x^{(13)} - 2.484 x^{(14)} + 0.791 x^{(15)} + 0.639 x^{(16)} + 0.425 x^{(17)} + 1.845 x^{(18)} + 0.558 x^{(19)}$$

Restricted Model

$$Y = 15.330u + 0.459 x^{(2)}$$

Where

Y = posttest scores

u = unit vector

$x^{(2)}$ = intelligence

$x^{(12)}$ = 1 if subject female in treatment group W, 0 otherwise

$x^{(13)}$ = 1 if subject female in treatment group X, 0 otherwise

$x^{(14)}$ = 1 if subject female in treatment group Y, 0 otherwise

$x^{(15)}$ = 1 if subject female in treatment group Z, 0 otherwise

$x^{(16)}$ = 1 if subject male in treatment group W, 0 otherwise

$x^{(17)}$ = 1 if subject male in treatment group X, 0 otherwise

$x^{(18)}$ = 1 if subject male in treatment group Y, 0 otherwise

$x^{(19)}$ = 1 if subject male in treatment group Z, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.427 \quad R_R^2 = 0.401 \quad F = 1.6821 \text{ NS}$$

Hypothesis XIII: There are no significant differences among

the treatment groups on posttest performance
when differences among them on pretest
have been controlled.

The regression equations expressing this hypothesis were:

Full Model

$$Y = 10.598u + 0.956 x^{(3)} + 2.075 x^{(8)} - 0.525 x^{(9)} \\ + 0.000 x^{(10)} + 0.809 x^{(11)}$$

Restricted Model

$$Y = 11.456u + 0.946 x^{(3)}$$

When

Y = posttest scores

u = unit vector

$x^{(3)}$ = pretest scores

$x^{(8)}$ = 1 if subject in treatment group W, 0 otherwise

$x^{(9)}$ = 1 if subject in treatment group X, 0 otherwise

$x^{(10)}$ = 1 if subject in treatment group Y, 0 otherwise

$x^{(11)}$ = 1 if subject in treatment group Z, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.547 \quad R_R^2 = 0.533 \quad F = 1.2434 \text{ N.S.}$$

Hypothesis XIV: There is no significant difference

between the two sexes on posttest per-
formance when differences between them
on pretest have been controlled.

The regression equations expressing this hypothesis were:

Full Model

$$Y = 10.013u + 0.967 x^{(3)} + 1.819 x^{(6)} + 0.000 x^{(7)}$$

Restricted Model

$$Y = 11.456u + 0.946 x^{(3)}$$

Where

Y = posttest scores

u = unit vector

 $x^{(3)}$ = pretest scores $x^{(6)}$ = 1 if subject is female, 0 otherwise $x^{(7)}$ = 1 if subject is male, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.545 \quad R_R^2 = 0.533 \quad F = 3.1819 (0.077) \text{ NS}$$

Hypothesis XV: There are no significant differences in

average posttest score among the sex by

treatment subgroups when the differences

among them on average pretest scores have

been controlled.

The regression equations expressing this hypothesis were:

Full Model

$$\begin{aligned} Y = & 10.291u + 0.982 x^{(3)} + 3.653 x^{(12)} + 0.589 x^{(13)} \\ & - 0.877 x^{(14)} + 1.591 x^{(15)} + 0.000 x^{(16)} - \\ & 2.958 x^{(17)} + 0.235 x^{(18)} - 0.391 x^{(19)} \end{aligned}$$

Restricted Model

$$Y = 11.456u + 0.946 x^{(3)}$$

Where

Y = posttest scores

u = unit vector

 $x^{(3)}$ = pretest scores $x^{(12)}$ = 1 if subject female in treatment group W, 0 otherwise

$x^{(13)} = 1$ if subject female in treatment group X, 0 otherwise

$x^{(14)} = 1$ if subject female in treatment group Y, 0 otherwise

$x^{(15)} = 1$ if subject female in treatment group Z, 0 otherwise

$x^{(16)} = 1$ if subject male in treatment group W, 0 otherwise

$x^{(17)} = 1$ if subject male in treatment group X, 0 otherwise

$x^{(18)} = 1$ if subject male in treatment group Y, 0 otherwise

$x^{(19)} = 1$ if subject male in treatment group Z, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.576 \quad R_R^2 = 0.533 \quad F = 3.7991 \quad 0.01225 \text{ Sign.}$$

Hypothesis XVI: There are no significant differences among treatment groups on posttest performance when differences among them on intelligence and pretest have been controlled.

The equations expressing this hypothesis were:

Full Model

$$Y = -10.374u + 0.241 x^{(2)} + 0.727 x^{(3)} + 1.639 x^{(8)} \\ -0.150 x^{(9)} + 0.000 x^{(10)} + 0.881 x^{(11)}$$

Restricted Model

$$Y = -10.249u + 0.249 x^{(2)} + 0.712 x^{(3)}$$

Where Y = posttest scores

u = unit vector

$x^{(2)}$ = intelligence

$x^{(3)}$ = pretest scores

$x^{(8)}$ = 1 if treatment W, 0 otherwise

$x^{(9)}$ = 1 if treatment X, 0 otherwise

$x^{(10)} = 1$ if treatment Y, 0 otherwise

$x^{(11)} = 1$ if treatment Z, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.630 \quad R_R^2 = 0.623 \quad F = 0.7811 \quad \text{NS}$$

Hypothesis XVII: There is no significant difference

between the two sexes on posttest performance when differences between them on intelligence and pretest have been controlled.

The equations expressing this hypothesis were:

Full Model

$$Y = -10.064u + 0.242 x^{(2)} + 0.725 x^{(3)} + 0.547 x^{(6)} + 0.000 x^{(7)}$$

Restricted Model

$$Y = -10.249u + 0.249 x^{(2)} + 0.712 x^{(3)}$$

Where Y = posttest scores

u = unit vector

$x^{(2)}$ = intelligence

$x^{(3)}$ = pretest score

$x^{(6)}$ = 1 if subject is female, 0 otherwise

$x^{(7)}$ = 1 if subject is male, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.624 \quad R_R^2 = 0.623 \quad F = 0.3723 \quad \text{NS}$$

Hypothesis XVIII: There are no significant differences in

average posttest scores among the sex by treatment subgroups when differences among them on average intelligence and pretest scores have been controlled.

The equations expressing this hypothesis were:

Full Model

$$Y = -8.620u + 0.231 x^{(2)} + 0.744 x^{(3)} + 1.214 x^{(12)} - 0.439 x^{(13)} - 2.025 x^{(14)} + 0.988 x^{(15)} + 0.000 x^{(16)} - 2.451 x^{(17)} + 0.000 x^{(18)} - 1.016 x^{(19)}$$

Restricted Model

$$Y = -10.249u + 0.249 x^{(2)} + 0.712 x^{(3)}$$

Where

Y = posttest scores

u = unit vector

$x^{(2)}$ = intelligence

$x^{(3)}$ = pretest scores

$x^{(12)}$ = 1 if subject female in treatment group W, 0 otherwise

$x^{(13)}$ = 1 if subject female in treatment group X, 0 otherwise

$x^{(14)}$ = 1 if subject female in treatment group Y, 0 otherwise

$x^{(15)}$ = 1 if subject female in treatment group Z, 0 otherwise

$x^{(16)}$ = 1 if subject male in treatment group W, 0 otherwise

$x^{(17)}$ = 1 if subject male in treatment group X, 0 otherwise

$x^{(18)}$ = 1 if subject male in treatment group Y, 0 otherwise

$x^{(19)}$ = 1 if subject male in treatment group Z, 0 otherwise

For this hypothesis the following were noted:

$$R_F^2 = 0.643 \quad R_R^2 = 0.623 \quad F = 2.0931 \quad NS$$

B29885